Content-Based Image Retrieval with Image Signatures

Nanayakkara Wasam Uluwitige Dinesha Chathurani
BScEng (Hons) 1st Class

A Thesis Submitted in Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy

at the

Queensland University of Technology
School of Electrical Engineering and Computer Science
Faculty of Science and Engineering

2017

Keywords

Information Retrieval
Image Decomposition
Image Signatures
Invariant
Normalisation
Feature Fusion
Membership Function
Robustness
Computational Efficiency
Random Indexing
Hierarchical Searching
Topsig
Clustering
Dimensionality Reduction
Bit Vectors
Relevance Feedback
Evaluation
Similarity Measure

Abstract

The development of the internet and the increased availability of image capturing devices have enabled collections of digital images to grow exponentially and become more diverse. This means more efficient and effective image browsing, searching, and retrieving tools are required by users in different fields such as education, medicine, economics, security, entertainment and architecture. Accurately retrieving images relevant to an information need from such large collections, is a challenging and important task to address. Over the last 30 years significant attention has been paid to content-based image retrieval. Extensive research has been conducted to develop advanced algorithms to extract low-level image features such as colour, shape, texture, edges, point of interest and spatial relationships, and to measure the similarity between pairs of images based on image feature vectors. Much work has been dedicated to exploring solutions to the problems of image rotation, translation and scale invariance. However, these algorithms cannot adequately model image semantics and have many limitations when dealing with broad content image databases, especially in regard to response time and retrieval accuracy.

Content-based image retrieval effectiveness vastly depends on the semantic gap between high-level semantic concepts used by people to understand image content and the low-level visual features extracted from images by automatic algorithms. Therefore, to develop an effective content-based image retrieval system, the semantic gap must be narrowed down. Narrowing down the semantic gap is an interesting and challenging task that motivates content based image retrieval researchers, and this research project. This research initially develops a method by combining low-level features such as shape, colour and texture. Though most

of the systems focus on specific datasets, this research addresses several general image databases which include colour images, texture images, and images with objects. Initially, different image features were studied with simple image retrieval techniques. Based on the knowledge acquired from the feature analysis, image features were selected. Images are represented using these features and in particular binary image signatures which represent images in binary format in lower dimensional space. These are then used for image retrieval to achieve better efficiency. The main objective of this method is to find suitable image blocks and adapt bag of features techniques analogous to the bag of words techniques used in text retrieval in the image retrieval context. Further, this research considers the dimensionality of feature vectors, which heavily influences the complexity of the similarity measure and reduces the dimension of feature vectors using random indexing. Experimental results were obtained for different evaluation measures on different standard general image collections. The results indicate the effectiveness of the proposed technique and show how the retrieval quality changes with the varying signature size which also influences the efficiency of the system.

Even though the system achieves good retrieval performance, it is desirable for the retrieval performance to be further improved in an effective way by narrowing the semantic gap. Pseudo relevance feedback has proven to be an effective mechanism for improving retrieval accuracy. Pseudo relevance feedback works by making the assumption that the images that were ranked highest in the initial search are relevant to the user. However, it does not take into account the relative ordering of results. To improve retrieval accuracy, a simple yet effective rank-based pseudo relevance feedback mechanism is proposed that weights the signal provided by an image considered relevant by the rank at which the image is retrieved. This pseudo relevance feedback mechanism works well with binary image signatures to improve retrieval precision.

Although the proposed pseudo relevance feedback approach improves the retrieval performance with the pseudo feedback and simulated feedback, better understanding of user feedback is desirable in order to improve the results by understanding the user viewpoint, which most systems have neglected to include. Therefore, the proposed rank-based relevance feedback approach was extended for

real user feedback. The experimental results highlight that interactive relevance feedback is important for improving the effectiveness of retrieval, and the user's viewpoint is different from classification assessments. Moreover, this approach provides a mechanism for end users to refine their image queries where users have no effective way to reformulate such an image query.

Scalability is also an important factor to consider in content-based image retrieval. The system's scalability is achieved in various ways. In this research, scalability is measured in the context of effectiveness, efficiency and robustness, which we defined in the thesis. Experiments were carried out to study these three factors; effectiveness was the main concern in all the proposed approaches and the effectiveness of each of the proposed approaches is provided in the evaluation section. The effectiveness of the overall system was considered, as well as the effectiveness of the system for different image collections. All these proposed approaches were evaluated for efficiency and the results show that the proposed approaches are efficient in retrieving images in milliseconds. Moreover, the system was evaluated for a range of image alterations to show the robustness of the system. Experimental results indicate the scalability of the proposed approaches.

The original contributions of this thesis can be further developed to increase the performance of all the system's aspects such as retrieval quality, speed and robustness.

Table of Contents

List of Figures

List of Tables

List of Publications

Publications Resulting from Research

The following fully-reviewed publications have been produced as a result of the work outlined in this dissertation.

International Conference Publications

The conference articles that have been published and submitted as part of this research are as follows:

(i) **N. W. U. D. Chathurani**, S. Geva, V. Chandran, and T. Chappell, "Content Based Image Retrieval Using Signature Representation," in *12th Australasian Data Mining Conference (AusDM14)*, 2014.

(ii) **N. W. U. D. Chathurani**, S. Geva and V. Chandran, "An Effective Content Based Image Retrieval System Based on Global Representation and Multi-Level Searching," in *10th IEEE International Conference on Industrial and Information Systems (ICIIS)*, 2015.

(iii) **N. W. U. D. Chathurani**, S. Geva and V. Chandran, "Content-Base Image (Object) Retrieval with Rotational Invariant Bag-of-Visual Words Representation," in *10th IEEE International Conference on Industrial and Information Systems (ICIIS)*, 2015.

(iv) **N. W. U. D. Chathurani**, S. Geva, V. Chandran and R.W.V.P.C. Rajapaksha, "Image Retrieval Based on Multi-Feature Fusion for Heterogeneous Image Databases," in *18th International Conference on Image Analysis and Processing (ICIAP)*, 2016.

(v) **N. W. U. D. Chathurani**, T. Chappell, S. Geva and V. Chandran, "Improving Retrieval Quality Using Pseudo Relevance Feedback in Content-Based Image

Retrieval," in *39th ACM International Special Interest Group on Information Retrieval (SIGIR)*, 2016.

(vi) **N. W. U. D. Chathurani**, S. Geva, G. Zuccon, V. Chandran, and T. Chappell, "Effective User Relevance Feedback for Image Retrieval with Image Signatures," in *21st Australasian Document Computing Symposium (ADCS)*, 2016. (accepted).

International Journal Publications

The journal articles that have been published as part of this research are as follows:

(vii) **N. W. U. D. Chathurani**, S. Geva and V. Chandran, "Conversion of an Image to a Document Using Grid-based Decomposition For Efficient Content-Based Image Retrieval," in *International Journal of Information Science and Intelligent System (IJISIS)*, Vol. 4, No.4, pp. 29–49, 2015.

List of Abbreviations

CBIR	Content Based Image Retrieval
ISIG	Image Signatures
RI	Random Indexing
TF−IDF	Term Frequency - Inverse Document Frequency
BoW	Bag of Words
BoF	Bag of Features
NN	Nearest Neighbour
SVM	Support Vector Machine
SIF	semantic image feature
FB	Feedback
RF	Relevance Feedback
PRF	Pseudo Relevance Feedback
SIM-RF	simulated user feedback
URF	real user feedback
RB-RF	Rank-Based Relevance Feedback
CH	Colour Histogram
CM	Colour Moment
CCV	Colour Coherence Vector
DWT	Discrete Wavelet Transform
IM	Image Moments
GFD	Generic Fourier Descriptor
EHD	Edge Histogram Descriptor

SIFT	Scale Invariant Feauture Transform
AP	Average Precision
ANMRR	Average Normalized Modified Retrieval Rank

Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signature:

Date: 11/11/2016

To my family

Acknowledgements

My experience as a PhD student at QUT has been more than pleasurable, which would not be the same without the support of my supervisors, colleagues, friends and family. Now that I am at the end of this PhD candidature, it is with much pleasure that I thank everyone who has helped me along the way.

Foremost, I would like to express my heartfelt gratitude to my loving parents who have been and still are my biggest supporters. Their never-ending support, as well as their guidance, comfort, compassion and their positive influence have allowed me to achieve more than I ever could have imagined. I am really grateful to them for their unconditional support throughout my life.

I would like to express my sincere gratitude to my principle supervisor, Associate Professor Shlomo Geva and my associate supervisor Professor Vinod Chandran for the continuous support given during my PhD study and also for their patience, motivation, enthusiasm and immense knowledge. Their guidance helped me through all the times I spent researching and writing this thesis. I would like to thank Professor Wageesh Boles for being an associate supervisor in the beginning of my PhD. I would like to thank my associate supervisor Dr.Guido Zuccon for the given support, reviews and the feedback later in my research journey. Without their encouragement and effort, my research and thus this thesis, would not have been completed.

My heartfelt thanks goes to the QUT high-performance computing (HPC) group, especially Mr.David Warne for the support related to HPC operations. I also express my sincere thanks to the staff members of the QUT Science and Engineering Faculty research office of QUT and EECS administration team, Ms.Janelle Fenner, Ms.Joanne Reaves, Ms.Joanne Kelly, Ms.Ellainne Steele and especially Miss.Judy Liu for their support in creating a productive research environment.

Further, I acknowledge Ms. Susan Gasson for her invaluable support. Also, I appreciatively acknowledge the financial support provided by QUT and the Australian government through out my PhD by the International Postgraduate Research Scholarship (IPRS) and the Australian Postgraduate Award (APA). I would also like to thank all academic and non-academic staff for the support given to me in countless ways. I would like to thank Amanda Greenslade who proofread the thesis word by word and corrected in order to improve my thesis.

I pay my sincere gratitude to my lab-mate Dr.Timothy Chappell, for helping me all the way during my PhD. Moreover, I would like to thank my lab-mates ,Lance De Vine and Dr.Inma Tomeos, for their support during my candidature. Also, I would like to express my warm and sincere gratitude to Medhavi Prasanthika, Dr.Madara Karunaratne, Dr.Dhanushka Krishnajith, Palmo Thinley, Dr.Rajitha Navarathne, Dr.Rupika Bandara, Dr.Sandya Wasanthi, Dilesha Senevirathna, Sarasi Madushika, Akila Pemasiri and Akmal Jahan and their families for their constant support and invaluable help. Their support and help was extremely important to me during my studied and is still now.

I would also like to thank my sisters and brother for the support and laughter you have given me over the years. Your encouragement and support always makes me a better person. I would like to thank my niece and nephew for giving me happiness so I could forget all the sorrows. I would like to give my special thanks to my loving friend, Dr.Cynthujah Vivekanathan, for everything you have done for me all these years until now. I would like to express my gratitude to my in-laws, all my relatives, my school and university teachers, my friends and those who have helped me directly or indirectly in this journey. Last but not least, I would also like to offer thanks from the depths of my heart to my husband, Kasun Kaluthanthri, for being by my side through the good times as well as the bad and encouraging me throughout this journey. You are only second to my parents.

N. W. U. D. Chathurani

Queensland University of Technology (QUT)

August 2016

Chapter 1

Introduction

Chapter Organisation

In the introductory chapter of this thesis, the motivation for the research to be presented is detailed in Section 1.1, followed by the research objectives and questions in Sections 1.2 and 1.3 respectively. Section 1.4 outlines the research outcomes and their underlying significance. The thesis organisation is summarised in Section 1.5.

The number of digital images is growing exponentially with image capturing for several reasons, e.g. for personal use (e.g., lifelogging [Gurrin et al., 2014, Hwang and Oh, 2015]), for security purposes [Yuan et al., 2011a, Zhang and Ye, 2009], etc. In addition, these images are often published on the web or shared within social network applications with the development of the internet, thus further enlarging the pool of images to which users have access. These trends have increased the demand for effective and efficient image searching methods which retrieve images according to their semantic meaning. Content-Based Image Retrieval (CBIR) was introduced to address these issues by automatically extracting image content to be used in the search process. CBIR methods allow for queries to be specified visually, e.g. through query-by-example methods. In CBIR, image features such as colours, textures, information shapes, etc, are extracted and indexed automatically with the aim to support the retrieval of images in answer to visual queries. Therefore, a number of CBIR systems were generated. However, each had their own drawbacks which require mitigation. Moreover, answering such queries, is still a rather

challenging task due to the semantic gap that exists between the low-level visual features of images and the semantics that humans associate to objects, entities and scenes in images. Different users may have different viewpoints about the same set of images. Thus, Relevance Feedback (RF) was used in some CBIR systems. It is an interactive process which is used to improve the retrieval performance by using user feedback (FB) while bridging the semantic gap. Figure 1.1 and figure 1.2 demonstrate the MARS [Mehrotra et al., 1997,Rui et al., 1997] CBIR system with RF.

In this system, the user is allowed to select an image as a query to retrieve similar images by using image features and the user is allowed to inform the system which of the retrieved results are relevant by providing preference weight for each relevant image as feedback. However, it may burden the user and sometimes confuse them because the user may only know that image is relevant but cannot judge how much relevant the image actually to the query. Though CBIR systems have been developed, they are far from satisfactory due to issues with user satisfaction, computational complexity and semantic retrieval, such as similarities between images that are subjective (changes in the viewpoint) and feature dependent.

Therefore, in this research, attention is given to developing a new CBIR system to address these problems.

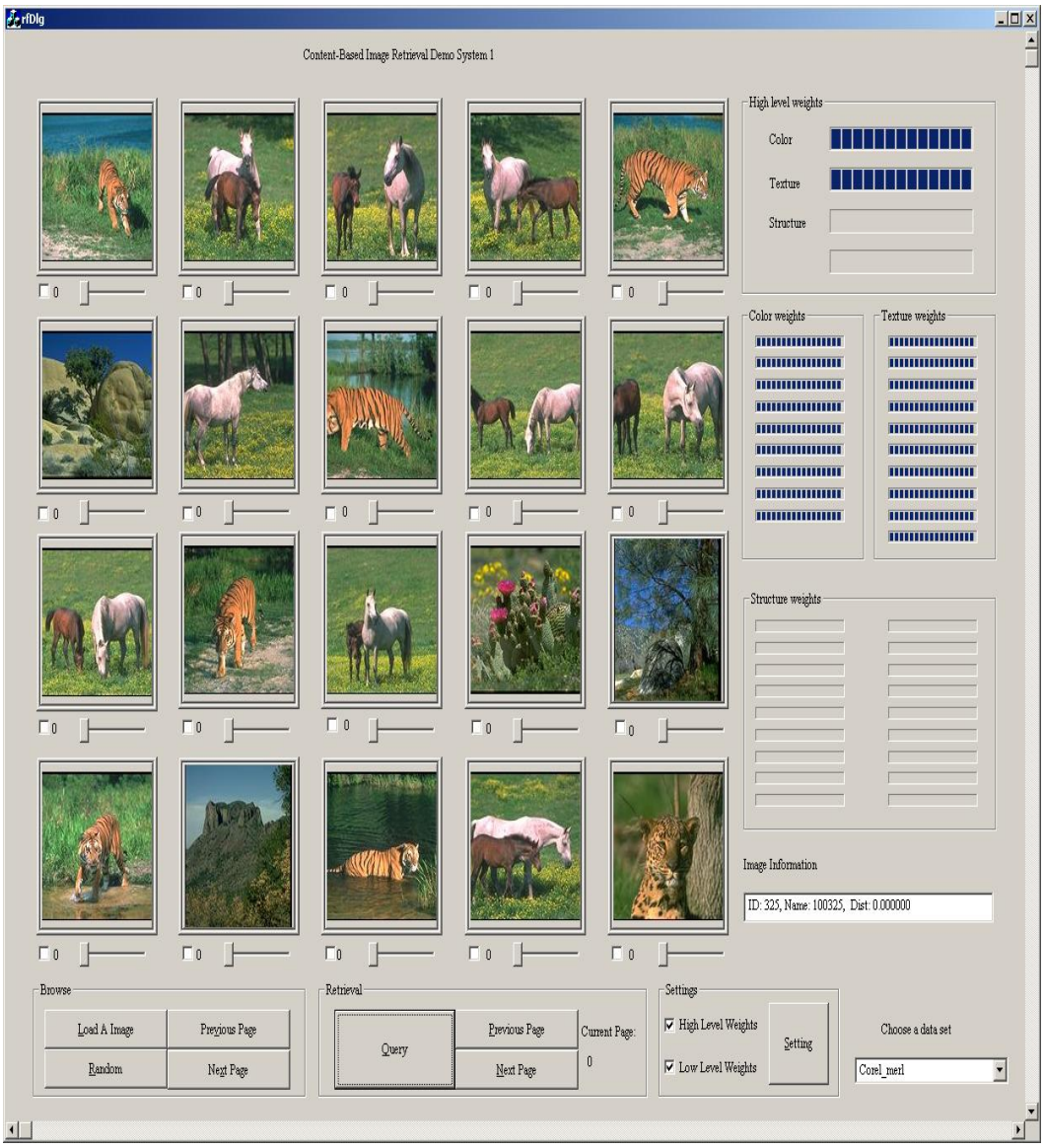


Figure 1.1: Retrieval results in the MARS system [Mehrotra et al., 1997, Rui et al., 1997]. The query image is the top left image.

Figure 1.2: Retrieval results in the MARS system [Mehrotra et al., 1997, Rui et al., 1997]. The query image is the top left image.

1.1 Motivation and Overview

When you are searching an image in the world's largest databases like Google and Bing, you may not be able to find images that match your expectation. Most image matches depend on annotated text that is associated to the image. Finding an image on a large database is a complex task due to the presence of thousands of irrelevant images to the query image. The next concept was to automatically extract the content of images for searching namely CBIR and that concept started in the 1990s and several techniques have been proposed to automatically extract the content of images. Moreover, many applications benefit from that, such as general image search, medical applications [Li et al., 2016], the video search [Jiang et al., 2016], surveillance [Nguyen et al., 2016], security [Zhang and Ye, 2009] and robotics [Lim et al., 2016].

General Image Search: People may want to search for images in, online large search engines, on a personal computer, or in a company database. In the first case, text-based image retrieval performance is poor due to several limitations, such as the inability to describe the content of the sought after image [Kato, 1992,Smeulders et al., 2000]. In the second and third cases, users normally cannot use text if those images are not tagged. Therefore, CBIR is used in these cases. For example, if I want to search some images in my image repository I can supply a similar image to the system and search. Then I can find images similar to what I search for.

Medical Applications: In the medical field, lots of images are generated every day, such as x-rays, ultra sound scans, molecular imaging and magnetic resonance imaging. It would be very useful if doctors could access similar cases and patients could access prior reports more quickly and easily with greater ease [Noteboom et al., 2014, Loorak et al., 2016].

Video Search: In this industry, lots of videos, films and advertisements are generated and stored in large archives which grow everyday. Therefore, it is difficult to find them when people need to access them. It would be very useful to provide a technique which could access them automatically by the content. Videos are generated by a collection of image frames, and images can be used to search a particular video when required.

Surveillance: The images captured by surveillance camera every second in different places which require high security and are stored. Those stored data are used to to detect suspicious people and unusual events automatically. Therefore, a content search is useful in order to understand the situation and the people.

Security: CBIR is useful in security purposes. It is used for facial recognition and fingerprint matching so that suspects can be identified. In private companies, they use these applications to verify their employers' identity to ensure security in the company.

Robotics: A machine or robot needs to be able to detect objects and situations. If the system has similar images, then the robot can identify the object by taking a picture and feeding that into the system.

During the last few years, CBIR techniques have been developed based on colour, texture, shape and spatial location. It is, however, still necessary to develop and solve problems related to segmentation, low-level feature extraction, representation, high-level semantics, storage and efficient indexing. Liu et.al. mention three levels of queries in CBIR in [Liu et al., 2007].

- Level 1: Retrieve images by low-level features such as colour, texture, shape or the spatial location of image elements. Example of such queries is 'find images with green background with brown object in the front', in other words 'find images like this'.
- Level 2: Retrieval of objects that are in the given query image identified by derived features. For example, 'find a picture of the Eiffel tower'.
- Level 3: Retrieval by abstract attributes, involving a significant amount of high-level reasoning about the purpose of the objects or scenes depicted. This includes retrieval of named events or activities, etc. Example of such queries is 'find pictures of a Sri-Lankan new year festival'.

We can see the most CBIR systems use level 1. However, levels 2 and 3 can be achieved by introducing interactive RF mechanisms to the system. Even though many CBIR systems are developed for image retrieval, describing the image according to the visual content is a challenging task for different reason. Two major reasons are the semantic gap and sensory gap. The sensory gap can be further subdivided into different categories, such as illumination changes, scale, viewpoint, occlusion and rotation.

	**Cigarette** **Tattoos** **Smoke** **Dress, Belt** **Sunglasses** **Shades**
	**Mailbox** **Fisheye** **Letter** **Self** **Portrait**
	**Iceland** **Ice** **Sky** **Clouds** **Lake** **Museum**
	**Village** **Trees** **Sky** **Evening**
	**Water** **Sight seeing** **Nature** **Island** **Downtown** **Mediterranean sea**

Figure 1.3: Semantic gap: The high-level concepts mentioned next to each box demonstrate the human vision and low-level features cannot significantly identify each and every aspect as user.

- **Semantic gap:** The semantic gap is the gap between the low-level features that are automatically extracted by machines and the high-level concepts of human vision and image understanding. Figure 1.3 provides an example for the semantic gap.
- **Sensory gap** The sensory gap is the gap between the object or the scene in the real world and its representation, which the system derives from the recorded image, such as illumination, scale and occlusion. They are briefly described as follows and shown in figure 1.4, figure 1.5, figure 1.6, figure 1.7, figure 1.8, figure 1.9 and figure 1.10.

  - **Illumination change:** Lighting conditions make important changes to a picture. Therefore, illumination changes affect the appearance. It is a trivial task to develop a robust system to identify an object or scene under different conditions of illumination.
  - **Scale/size change:** An object may appear at different scales in different images and the object's placement may be different from one to the other. Moreover, scenic images may appear up closer or far away, such as a set of trees at the front of an image, while the same trees may appear behind grassland in another image. These images which can be considered the same.
  - **Changes in viewpoint:** The position of the camera in relation to the object or scene can change its appearance. This may lead the system to recognise those images as different.
  - **Occlusion:** Occlusion happens in various ways. Occlusion occurs if the main object of an image is hidden (occluded) by another object. Sometimes occlusion is the areas we do not have any information about (part of an object).
  - **Rotation:** There may be cases of images with an object or scene that have different angle. They may be rotated clockwise or anticlockwise, or be upside down. Even though the object or the scene is identical except for the angle, the system may consider them different.
  - **Other:** Background clutter, truncation and articulation.

Figure 1.4: Illumination changes.

Figure 1.5: Scale/size variation.

Figure 1.6: Changes in viewpoint.

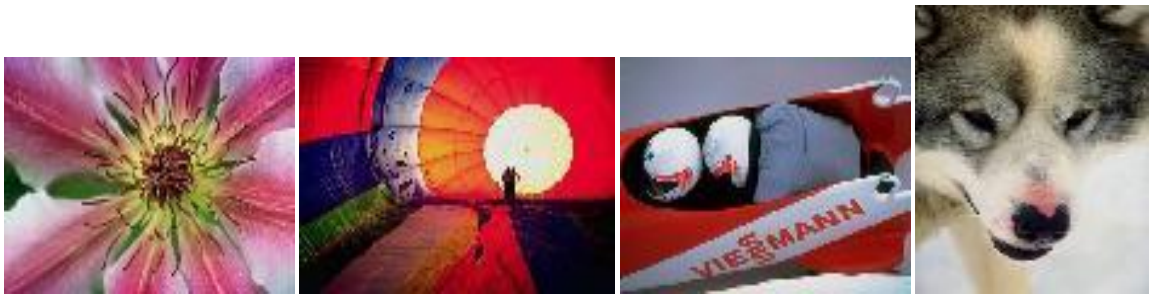


Figure 1.7: Occlusion and truncation.

However, a human can handle all these situations, unlike CBIR systems. Apart

Figure 1.8: Rotation variation.

Figure 1.9: Background clutter.

Figure 1.10: Articulation.

from these there are other factors, such as ambiguity and the viewpoint of users, which affect the performance and cause problems of incomplete query specification or incomplete image description. User viewpoint may be totally different from the classification. Therefore, real user feedback is effective for performance improvement in CBIR. Normal evaluation using groundtruth, PRF and simulated user RF can be used to evaluate the classification performance and real users are essential to the evaluation retrieval performance. However, intra-class variability and inter-class variability effect the classification performance.

If the system is capable of bridging the semantic gap and allowing the queries

to be satisfied, then it will be a smart system, as per the users' expectation. But it is a complex and challenging task and thus, it is not yet solved.

While bridging the semantic gap, it is essential to consider the other important issue, which is the retrieval performance. It is true that users normally are not satisfied with a slow system which takes a considerable amount of time to give results. However, system search time is affected by the curse of dimensionality and the large number of objects to search through. So indexing and retrieval in a CBIR system must be efficient with low computational cost. Memory and disk space requirements also affect the overall performance of the system. The usability of the system also needs to be considered. So it is better to address this issue when developing a CBIR system. If a system can satisfy the aforementioned issues, especially semantic retrieval performance, that system will be very effective, but is a challenging task, as mentioned above.

1.2 Research Objectives

Motivated by the reasons stated in the previous section, the research presented in this thesis has the following aims and objectives:

The primary aim of this research project is to develop an efficient and effective general purpose CBIR system with image signatures.

The objectives of this research project are to

(*i*) Study existing CBIR systems and identify their advantages and limitations and find an effective combination of features that captures content information from images for CBIR.

(*ii*) Develop an image signature-based CBIR system by using techniques analogous to text retrieval to address large data collections.

(*iii*) Discover the effect of pseudo relevance feedback and interactive user relevance feedback and evaluate the system's performance.

(*iv*) Investigate the scalability of the developed system and relevance feedback approaches.

1.3 Research Questions

The specific research questions that have been addressed in this research are as follows:

> *How can we achieve better retrieval performance in ways analogous to text retrieval by using Bag of features (BoF) and Random Indexing (RI)?*

This question entails an investigation aimed at identifying the image features and the way of representing images.

The following sub-questions arise in relation to image search using BoF:

(*a*) *Which low-level features can be used effectively to extract image features?*

(*b*) *How to extract those features from images?*

(*c*) *What kind of image representation can be derived from low-level features to facilitate an effective search?*

Once these questions have been answered, a CBIR system can be generated by using BoF approach to achieve better retrieval performance. The next target is to improve the retrieval performance further. This leads us to the next research question:

> *Could RF on images be applied in the same way as signature-based RF on text documents to improve retrieval performance?*

This question addresses how can we improve CBIR performance using an RF approach analogous to text. Then we can find a technique that can be applied on CBIR with RF. Further, the suitability of this technique for image retrieval can be verified using a set of design and evaluation criteria. Those criteria are to be based on how best it can satisfy the users' image requirements. The following sub-questions arise in relation to achieving an efficient CBIR system:

(*a*) *Does the rank have an impact on relevance feedback ?*

(*b*) *Which rank-based relevance feedback method is useful?*

(*c*) *How can we use pseudo relevance feedback?*

- *How can we handle relevance feedback with binary image signature?*

(*d*) *What effect does interactive relevance feedback have on image retrieval system?*

- *How to evaluate the relevance feedback?*

After the RF scheme is introduced we must consider the user's viewpoint. Then next question arises because the user always expects a fast system:

> *How can we achieve fast retrieval results and scale this approach to a very large image database?*

This question entails an investigation aimed at providing efficient indexing and retrieval by having low computational cost. The following sub-questions arise in relation to achieving an efficient CBIR system:

(*a*) *Which representation will be the most useful to save storage space and time?*

(*b*) *Which search mechanism is suitable?*

(*c*) *How does the searching time vary with the size of the dataset?*

1.4 Research Contributions

In addressing the research objectives and questions, this thesis presents a number of original contributions and achievements in the field of CBIR systems. The main contributions are summarised below.

(*i*) Simple image retrieval systems were developed with global and local image representations to achieve better retrieval accuracy as preliminary work (Chapter 4 and Publication (ii and iv) in the List of Publications).

- Initially, the suitable features were selected and a CBIR system based on global representation and multi-level searching was developed. This proposed system is unique, as it considers one feature at each step and uses the results of the prior step as input for the next step in a multilevel manner, whereas, in past methods, all the features were fused at once for the single-level search of a typical CBIR system. The

proposed approach is simple and easy to adopt. The retrieval quality of the proposed approach was evaluated using two benchmark datasets for image classification. The proposed system showed good results in terms of improvement in retrieval quality in comparison with the literature.

- Then another CBIR approach was developed with a multi-feature fusion approach for efficient CBIR, based on the distance distribution of features and relative feature weights at the time of query processing. It is a simple yet effective approach, which is free from the effect of the features' dimensions, ranges, internal feature normalisation and distance measure. This approach can easily be adopted to any feature combination to improve retrieval quality. The proposed approach was empirically evaluated using two benchmark datasets for image classification and was compared with existing approaches. The performance of the proposed approach is demonstrated with the improved performance in comparison with the independently evaluated baseline of the previously proposed late-feature fusion approaches.
- The original contribution of the first method is the multi-level search mechanism in the context of CBIR systems. Previous methods [Li et al., 2000, Chen and Wang, 2002, Takala et al., 2005, Hiremath and Pujari, 2007a,Hiremath and Pujari, 2008,Yuan et al., 2011b,Saad et al., 2011, Mansoori et al., 2013] typically use a single level search after feature fusion. Furthermore, this provides higher retrieval performance than single-level search. The second method, the late-feature fusion method provides robustness against changes in feature dimension, range of values, normalization and distance measures and is shown to provide better performance than existing late-feature fusion methods.

(*ii*) A CBIR system was developed based on a novel image representation using a new approach to the generation of image signatures namely "Content Based Image Retrieval with Image SIGnatures"(CBIR-ISIG) (Chapter 5 and Publication (i, iii, and vii) in the List of Publications)

  - Initially, semantic image blocks that can be used to extract features were identified, then features were extracted and vocabularies were generated. Each block was characterised by colour, texture and shape features using BoW with the feature index and the index of the nearest cluster centre. Then image signatures were generated by applying RI to a Bag-of visual Words (BoW) representation of the images. Then the CBIR-ISIG system was developed with these signatures and the performance of the proposed approach was evaluated using three benchmark datasets for quality, speed and robustness. This results showed this approach has the high potential to retrieve correct images and this can be extended to a large collection as the binary signatures increase the efficiency in searching. System performance was compared with existing systems in the literature and the results highlighted that our approach has a superior performance over the other systems.
  - Parameters were evaluated to select the most suitable parameters, and different evaluation measures were used to demonstrate the performance.
  - As an extension of the proposed CBIR-ISIG system, a Rotation Invariant Bag of visual Words (RIBoW) approach was introduced which encodes spatial information in order to achieve effective CBIR results, especially in object-based retrieval. The RIBoW approach uses circular image decomposition in combination with circular shift operation to achieve invariance in rotation using global image descriptors. The retrieval quality of the proposed approach was empirically evaluated using two benchmark datasets for image classification and the results were compared with existing systems in the literature. The performance of the proposed approach highlighted its effectiveness and robustness for rotation invariant image(object) retrieval by transcending retrieval performance when compared with BoW approaches.

- while BoW and RI are not new concepts, in this research we adapted them in a novel context by generating image signatures using RI as explained above. Previous methods instead commonly have used semantic analysis,

singular value decomposition and locality sensitive hashing. Moreover, our approach provides effective and efficient image retrieval compared to other available systems. It also provides the opportunity to prefer effectiveness over efficiency or vice versa.

(*iii*) A relevance feedback approach; a "Rank-Based Pseudo Relevance Feedback" (RB-PRF) approach was developed to improve retrieval performance through the incorporation of feedback in response to an initial results list (Chapter 6 and Publication (v) in the List of Publications)

- A novel rank-based pseudo relevance feedback approach was introduced which explicitly incorporates the original rank of the results and was used as implicit relevance indicators. Therefore, this deviates from the common use of PRF that ignores the rank positions of the feedback documents. To integrate the rank in RF, a scaling factor was introduced after experimentation with several scaling factors.
- This RB-PRF mechanism innovates by making use of binary image signatures to improve retrieval precision by promoting images similar to highly-ranked images and demoting images similar to lower ranked images. This RB-PRF approach uses document signatures directly in feedback processing, as opposed to traditional approaches, which return to the original images (or documents). The top-ranked signatures are already resident in memory, thus there is no need to work with the original documents at run time. The use of signatures and the allocation of resident memory for their storage allows for an extremely efficient retrieval method, both in terms of memory usage and runtime.
- All the parameters were evaluated and explored the role of each parameter on the effectiveness of RB-PRF approach using different datasets. Empirical evaluations based on standard benchmarks demonstrated the effectiveness of the proposed RB-PRF mechanism in image retrieval.

- Even though PRF was used in CBIR in earlier systems, they have neglected the rank positions of images provided as feedback. Our novel RB-PRF incorporates the original rank of the feedback images and as an implicit relevance

indicator to improve retrieval effectiveness. In addition, the application of RF directly on image signatures is novel: this approach saves time and simplifies the RF process.

(*iv*) The RB-PRF experiments were extended using simulated and real users to achieve better performance and to understand the user intention (Chapter 7 and Publication (vi) in the List of Publications).

- Experiments were carried out with real users to demonstrate how to use explicit relevance feedback effectively with signature-based image retrieval in order to improve retrieval quality.
- This approach provides a mechanism for end users to refine their image queries. Unlike text retrieval systems where users are able, and generally prefer, to reformulate their text queries to improve search results, there is no effective way to reformulate an image query. This approach provides a solution to this problem.

- Mainly the binary signature based CBIR system with RF is efficient, faster and scalable for very large databases and robust to range of image degradations. Then RB-RF further improves retrieval performance efficiently and provides a mechanism for users to refine their queries easily. This Topsig based binary signature representation and how it is generated, rank-based approach and how it is achieved is new to CBIR.

1.5 Thesis Organization

The remainder of this thesis is organised as follows:

Chapter 2 captures the background of the proposed research of this thesis, with an overview of CBIR systems. Apart from specific details on CBIR systems, relevant concepts of CBIR systems are reviewed. Challenges in CBIR are analysed, paying special attention to those addressed in this dissertation. While a comprehensive literature review on CBIR is conducted in this chapter, a more specific literature review on each of the key research topics is provided in each chapter of this dissertation.

Chapter 3 presents, in detail, the different databases and systems, evaluation measures, evaluation methodologies and parameter settings used to evaluate the CBIR methodologies laid out in this thesis.

Chapter 4 presents the preliminary works which were done in this research. Initially, it describes the image descriptors and colour spaces which were selected in the beginning for this research. Then an effective CBIR system based on multilevel searching is proposed and evaluated using several datasets to demonstrate the retrieval performance. Finally, a simple image retrieval mechanism based on late feature fusion is proposed and the evaluation is discussed to show the retrieval quality.

Chapter 5 presents the design, development and performance evaluation of the proposed signature-based image retrieval system (CBIR-ISIG). An overview of the design of the CBIR-ISIG system is first provided. The process of each step in the CBIR-ISIG system is then explained. The performance of the CBIR-ISG depends on several factors, namely feature selection, representation, the size of the semantic image block, and the signature size. The effect of all these factors is analysed in detail. Different databases are used to demonstrate the effectiveness of the proposed technique regardless of the particular database used. Finally, an application of the image signatures to the CBIR is introduced.

Chapter 6 presents the design, development and performance evaluation of the proposed rank-based pseudo relevance feedback (RB-PRF) approach. The first pass retrieval of CBIR-ISIG is discussed in the beginning. Then the RB-PRF approach is explained with its paramter setting. The performance of the RB-PRF retrieval depends on several factors, namely the scaling factor, the sample size, the list size, and the database size. The effect of all these factors is analysed in detail. Three different databases are used to demonstrate the effectiveness of the proposed technique regardless of the particular database used.

Chapter 7 presents an extension of the RB-PRF approach proposed in Chapter 6 to the real user. This chapter studies the real user feedback and the performance was evaluated using simulated feedback and real user feedback.

Different evaluation measures were used to compare the performance with existing mechanisms and self-comparison without feedback and with PRF.

Chapter 8 discusses the scalability of the proposed CBIR-ISIG system. Initially, it discusses the effectiveness of the system to retrieve correct images followed by the efficacy of the system to retrieve images. Then the usability of the system is explained, as it is a vital factor when dealing with the real users. Finally, the robustness of the system is presented.

Chapter 9 concludes the thesis with a summary of the original contributions and future work.

Chapter 2

Background

Chapter Organisation

This chapter begins with a history of the evolution, concepts and general challenges in image retrieval in Section 2.1, 2.2 and 2.3 respectively. Section 2.4 includes information about existing Content-Based Image Retrieval (CBIR) systems, their features and their influence. The most popular and most-cited datasets are presented in Section 2.5.1. A comprehensive overview of the visual features that can be extracted from images and the feature selection mechanisms and different image representation methods are discussed in Section 2.5.2. Section 2.5.4 presents the dimensionality reduction methods followed by indexing mechanisms. Then Section 2.5.5 and Section 2.5.6 describe the user's information needs and the results visualisation of CBIR systems. A comprehensive overview of the Relevance Feedback (RF) mechanism which is used to reduce the semantic gap is given in Section 2.6. The chapter's summary and conclusions are included in Section 2.7.

2.1 Evolution

The development of the internet and increased availability of image capturing devices have enabled collections of digital images to grow at a fast pace in recent years and to become more diverse. This has created an ever-growing need for efficient and effective image browsing, searching and retrieval tools. Retrieving

relevant images accurately to satisfy an information need from a large, diversified collection is a challenging task. Therefore, the field-of-image retrieval gained interest and was first started in the late 1970s [Palermo and Weller, 1980] and has been a very active research area.

The early system was Text-Based Image Retrieval (TBIR) and the basic annotation method was through associated keywords. Initially, images were annotated by text and a database management system was used to retrieve images based on text, e.g. the words used as a caption to an image in an article or acquired through manual annotation processes (e.g. tag-based annotations) [Chang and Fu, 1980,Rui et al., 1999]. However, this method had a limited performance, especially on large image sets. The main problems of TBIR methods are:

- The automatic generation of keywords (annotation) for a large collection is not yet feasible. Therefore, a huge amount of human effort is required to manually annotate the images. Thus, it is a very expensive and cumbersome task, especially when the size of the dataset is large.
- Manually created logical image representations are highly subjective by nature, as the rich contents in images and different level of prior knowledge and experience can influence the understanding of an image.
- Annotations are context-sensitive. Different people may see different things in an image and the same person may see different things at different times.
- Manual annotations are often incomplete as it is difficult and cannot be clearly described in some features such as complex texture.
- When querying, performance may be slow due to the lack of syntax and can be very complex in order to generate queries.

Further studies to overcome the aforementioned problems introduced the concept of CBIR in the early 1990s [Kato, 1992, Smeulders et al., 2000] as a potential solution. CBIR is based on the visual content of images such as colour, texture and shape features. Since then, a lot of new techniques have been proposed to extract content characteristics automatically from the images, and to manage and use indexing to improve the CBIR performance. The development of the internet created an ever-growing need for efficient and effective image browsing, searching

and retrieval tools. Despite many years of research in CBIR, an effective general solution in terms of speed, precision and scalability is still challenging for researchers.

2.2 Concept

The main objective of CBIR is to find an image or set of images which are similar to the query image provided by the user. Figure 2.1 illustrates the concept of CBIR which contains the following components [de Brito Ferreira, 2010]:

- **Database:**

  Most image retrieval systems work with closed databases and are indexed offline. The features of each image on the database are extracted through the visual descriptors and generate a multidimensional feature vector. Then the feature vectors are stored. There are different kinds of visual descriptors available for feature extraction.

- **Information Need and Query Processing:**

  In here, the query image provided by the user is processed. The system uses Query By Example (QBE). This may be an image or sketch drawn by the user and it may be one or multiple images. The same process that was used to convert the image database to the internal representation of feature vectors is used to generate the representation for the query.

- **Result Generation and Presentation:**

  Results are generated by making a comparison between the feature vector of the query image/s and the feature vectors of the database images using suitable techniques. Then the images are ranked according to the similarity measure and the top-ranked images are displayed to the user.

- **Relevance Feedback:**

  Some systems allow the user to refine the retrieved result by providing RF. RF is the fine tuning of the query process which can be used improve results by reducing the semantic gap. The user provides feedback by indicating whether the shown images are relevant (positive example) or irrelevant (negative feedback) to the system. Based on the feedback information, its

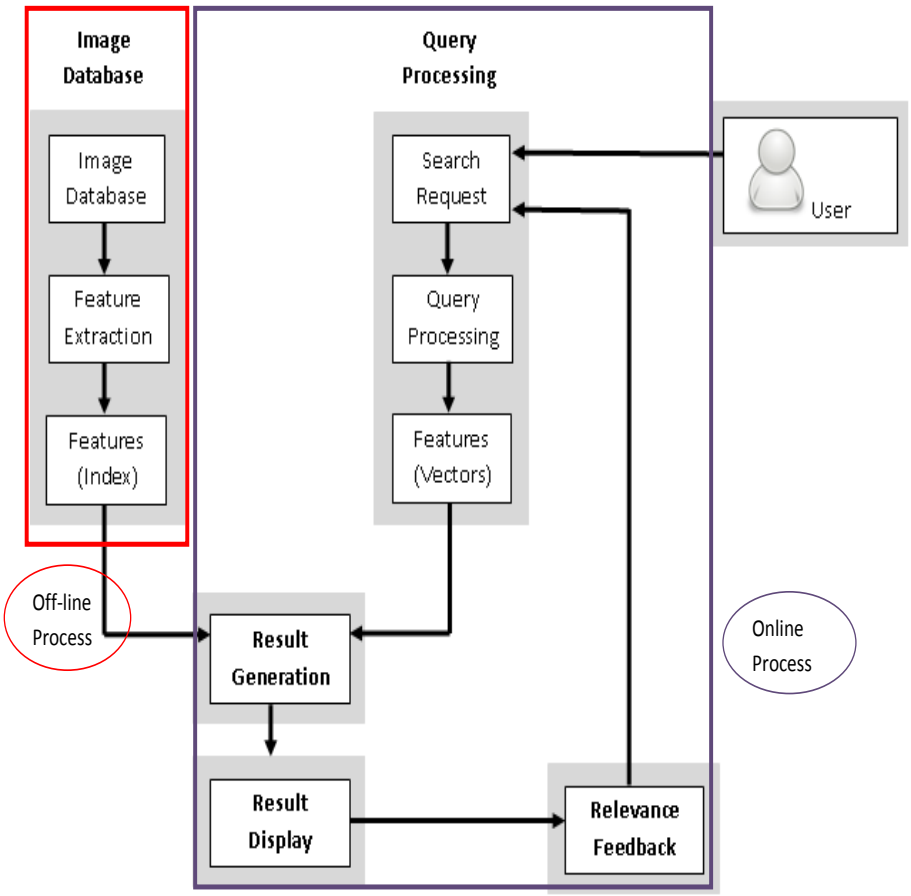


Figure 2.1: An overview of a CBIR system.

necessity is refined and starts the image retrieval process over and over again until the user is satisfied. It improves the retrieval performance and it is an optional step in a CBIR system.

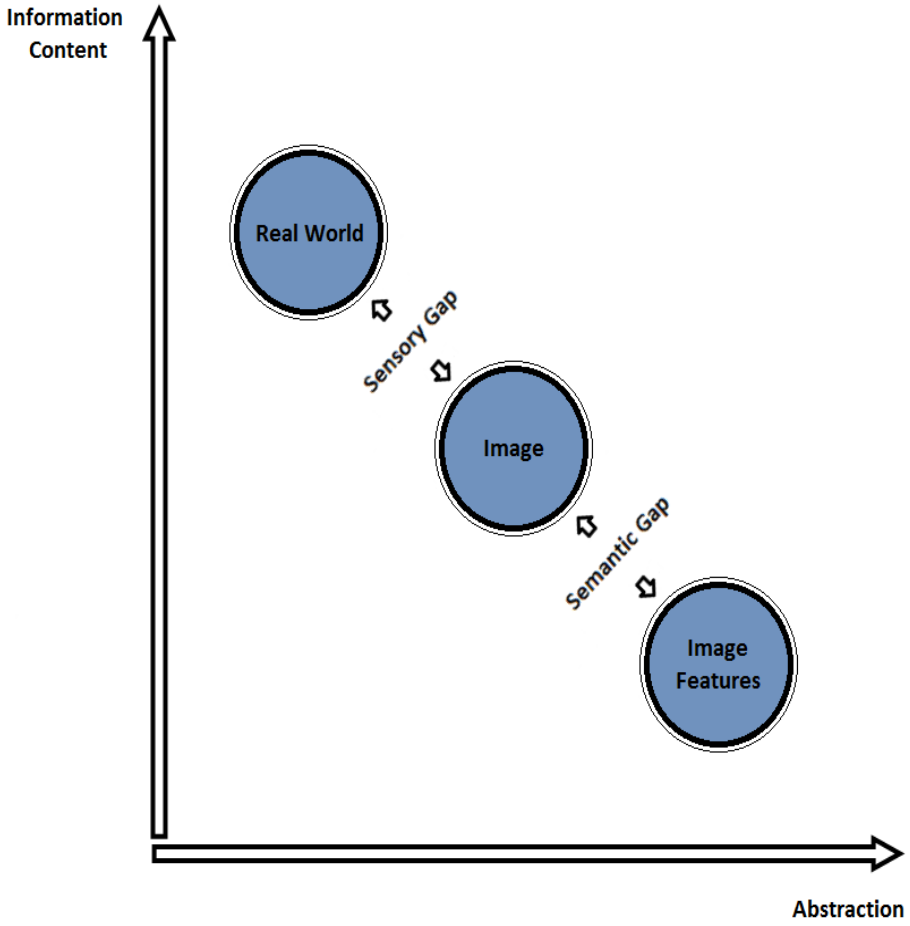


Figure 2.2: The sensory and semantic gap of an image [de Brito Ferreira, 2010].

2.3 Overview of Current Challenges

Even though CBIR solves the problems in TBIR and has its advantages, it also has many problems. The two most important challenges in CBIR are sensory gap and semantic gap as shown in figure 2.2.

- **Sensory gap:** To generate an image, it has to be captured by a device. When an image is produced, some of the information that was presented in the real world is automatically lost. This loss may be due to different factors, such as low resolution, bad illumination, partially occluded objects, the viewing angle or any fault in the capturing device (camera). The sensory

gap is defined as the "gap between the object in real world and the information (computational) description derived from recording of that scene (image information)" [Smeulders et al., 2000].

- **Semantic gap:** There is always a gap between low-level visual features extracted from an image in order to describe the image and the high-level semantic concepts (objects, relationships, meanings, feelings) required by the system users. The semantic gap is defined as the "lack of coincidence between the information that one can extract from visual data and the interpretation that the same data have for a user in a given situation" [Smeulders et al., 2000].

Some of the other problems encountered are:

- The query specification for CBIR, such as selecting colours, shapes or texture patterns that are very cumbersome and all the users may not aware of the image features.
- CBIR is more computationally complex than TBIR and takes considerable time to process and complete a search.
- The performance deteriorates if a suitable image cannot be used as the query.
- The image features are high-dimensional and affect the searching time, memory and disk space requirements.

More and more CBIR techniques have been introduced and still researching to provide solutions to these problems in image image retrieval. Articles [Smeulders et al., 2000], [Goodrum, 2000], [Lew et al., 2006], [Oussalah, 2008] and [Dharani and Aroquiaraj, 2013] provide excellent comprehensive reviews of CBIR, challenges and future directions. Some of the famous and mostly-cited CBIR systems are discussed in the next section.

2.4 Existing CBIR Systems

CBIR aims at developing techniques that support effective and efficient searching and browsing of images by generating semantically accurate results of large and highly-varied image datasets based on automatically extracted image features. Developing a universally-accepted CBIR is a challenging task because of the following:

- selecting a semantically precise image segmentation method to obtain accurate regions, finding the best-suited image features and feature extraction methodologies, representation, mapping low-level features to high-level semantics, image indexing and image similarity measuring using optimisation techniques and retrieval.

In general, CBIR is an active research area in which ambitious attempts have been made and yet so many problems still need to be solved to satisfy users' expectations. Although CBIR is still immature, a large number of general-purpose commercial and academic image retrieval systems have been developed in the past decades. Some of the most well-known and most-cited systems are IBM QBIC [Flickner et al., 1995, Niblack et al., 1993], MIT Photobook [Pentland et al., 1996] , VisualSEEK [Smith and fu Chang, 1996a, Smith and Chang, 1996, Smith and fu Chang, 1996b], NeTra [Ma and Manjunath, 1997], MARS [Mehrotra et al., 1997,Rui et al., 1997], PicHunter [Cox et al., 2000], MINDS [Takahashi et al., 2000], SIMPLIcity [Wang et al., 2001], PicSOM [Laaksonen et al., 2002], Blobworld [Carson et al., 2002, Belongie et al., 1998] and CLUE [Chen et al., 2003, Chen et al., 2005].

QBIC

The Query By Image Content system [Flickner et al., 1995, Niblack et al., 1993] has been developed by IBM. This is the first commercial CBIR system. This system depends on a number of features that can be selected by the user. It uses properties such as colour percentage, colour layout and texture for image-based similarity comparison. It allows users to submit query-by-example images, user constructed sketches and drawings and selected colour and texture patterns. The colour features used in the system are average R,G,B (R-Red, G-Green and B-Blue), Y,i,q(Y-Luminance, i-inphase, q-quadrature; luminance and chromaticity information), L,a,b(L-Luminance, a and b are the chromatic components) and a mathematical transform to Munsell(MTM), and represent in a colour histogram. The texture features used here are an improved version of Tamura texture representation. As a shape feature, they used shape area, circularity, eccentricity, major axis orientation and algebraic moment invariants. The dimensionality reduction techniques are (Kullback-Leibler or principal component analysis)

applied for high-dimensional features. The similarity measure is done by Euclidean distance (colour and texture) and Quadratic form distance (colour).

Photobook

Photobook [Pentland et al., 1996] has been developed at the Media Laboratory, at Massachusetts Institute of Technology (MIT), and it is one of the first academic prototype CBIR system. Queries are allowed as a text annotation on images. It is a set of interactive tools for searching and querying images. The system uses colour, texture and shape features for image retrieval. Texture features are extracted as Wold components which are namely periodicity, directionality, and randomness. Shape feature is described by extracting the shape boundary and then taking corners and curvature points. The main system is divided into three specialised systems Appearance Photobook, Texture Photobook and Shape Photobook, which can also be used in combination. The incorporation of an interactive learning agent named FourEyes for selecting and merging feature-based model is a unique feature of Photobook. This approach is found to be effective in interactive image annotation.

VisualSEEK

The VisualSEEK [Smith and fu Chang, 1996a,Smith and Chang, 1996,Smith and fu Chang, 1996b] search engine was developed at Columbia University, USA. The system uses colours, sizes and both absolute and relative arbitrary spatial layouts of colour regions for measuring similarity. HSV colour space is used. Colour set, location, area and spatial extent are also used as features. It supports queries based on both visual features and their spatial relationship. It uses automated region extraction and the features are extracted from compressed domains. A binary tree indexing algorithm has been used to speed up the process.

NeTra

NeTra [Ma and Manjunath, 1997] is a prototype image retrieval system that was developed at the University of California, Santa Barbara. It supports low-level features of colour, texture, shape and the spatial information of

segmented regions to search similar regions from the database. The colour feature is extracted by colour histogram, the shape is represented by the Fourier transform of contour representation, and Gabor filters are used for texture feature extraction. Images are segmented into regions using edge flow-based region segmentation that allows an object or region-based search. This representation allows users to submit queries based on features that combine regions of various images in the database such as "retrieve all images that contain regions having colour of object A, texture of object B, shape of object C, and lie in the upper one-third of the image" [Ma and Manjunath, 1997]. It uses Gabor filter based texture analysis and neural net-based image thesaurus construction.

MARS

The Multimedia Analysis and Retrieval System [Mehrotra et al., 1997, Rui et al., 1997] is an information-centric approach, which was developed by the University of Illinois. To represent images, it uses colour, texture, shape and layout features as well as a manual text description of the image. A colour histogram of HS values describes the colour and coarseness, contrast and directionality provides the texture details, while shape is described by a Fourier transform of the shape boundary. The user is allowed to select a combination of specific features from the selected images as the query. A Boolean (fuzzy and probabilistic) retrieval model is used for retrieval purposes. Term frequency, inverse document frequency measure (tf-idf) and RF is used together to achieve better performance.

PicHunter

PicHunter [Cox et al., 2000], is prototype CBIR system. It uses colour and colour spatial information for retrieval. It also uses a colour histogram of HSV values, a colour autocorrelogram of HSV, and a colour-coherence vector of RGB. It represents a simple instance of a general Bayesian framework. RF is used to improve the quality by reducing the semantic gap. In here, they use RF in a different manner. Unlike other systems, it concentrates the user's RF from the beginning (not only the feedback of the previous iteration) during retrieval.

MINDS

The Movie INDexing System [Takahashi et al., 2000] was developed by an imaging system business group at Ricoh Co. Ltd. MINDS is a prototype retrieval system which supports still images and movies. It uses the structural information of the image, such as spatial texture and the spatial edge feature. The texture feature is extracted by a Gray level co-occurrence matrix and the shape feature is extracted by an edge orientation histogram for four orientations. The system partitions an image into blocks and extracts the features from each.

SIMPLIcity

Semantic Sensitive Integrated Matching for Picture LIbrary [Wang et al., 2001]. This system classifies images into semantic categories such as textured/nontextured, and graph/nongraph. It incorporates an integrated region matching methodology that reduces the adverse effects of inaccurate segmentation. It supports low-level features of colour, texture, shape and the location of segmented regions.

PicSOM

The name stems from "Picture" and the Self-Organizing Map (SOM) [Laaksonen et al., 2002]. It is a neural network-based CBIR system. PicSOM uses a subset of MPEG-7 standard visual content descriptors. The colour feature is extracted in four ways: dominant colour- the first and second most dominant colour of the LUV space, scalable colour- the colour histogram in the HSV colour space, colour layout- the coefficient of the discrete cosine transform of dominant colours of the YCbCr space in 8*8 blocks, and colour layout. Texture is extracted as an edge orientation histogram, and angular radial transform of the shape region provides shape details. Tree Structures SOM (TS-SOM) are used to achieve a hierarchical structure and reduce the complexity of the search. This framework builds upon Query By Pictorial Examples (QBPE) and RF.

Blobworld

The Blobworld [Carson et al., 2002, Belongie et al., 1998] representation is

created by clustering pixels in a joint colour-texture-position feature space with a mixture of Gaussians using expectation-maximisation and the minimum description principle. It was developed by the computer science division of the University of California at Berkeley. In this system, when querying, the user is allowed to access the regions directly to specify which are important. Before feature extraction, the image is segmented into regions. It uses colour, texture and position features to describe an image. Eight features are used for segmentation. A three dimensional colour descriptor of CIE L*a*b* space is used as colour, mean contrast and anisotropy are used as texture and approximate area, and eccentricity and orientation are used as spatial descriptor. A Blobworld representation of each retrieved image is shown to the user.

CLUE

CLUster-based rEtrieval of images by unsupervised learning [Chen et al., 2003, Chen et al., 2005], retrieves image clusters by reducing the semantic gap. It uses a segmentation technique and the images are divided in to 4*4 blocks, and each one is described by colour, texture and shape. The average colour (in the LUV colour space) of the corresponding block is used as colour energy and L values of one-level Daubechies-4 wavelet transform is used as texture feature. The system generates clusters by using not only the feature similarity but also how images are similar to each other. One of the specialties of this system is that it can be embedded into a typical CBIR system without considering the features of the system, such as features used, whether RF is used or not, sorting method, etc.

All these CBIR systems were developed for different reasons and they have their advantages and weaknesses. Some of the weaknesses of these systems are, the user is asked to select the feature to be used in the retrieval process which require prior knowledge, uses text queries and image descriptors have been selected only covering one or two features. Moreover, almost all the systems were not scalable for very large databases.

Recently neural networks and Convolutional Neural Networks (often referred to as deep learning) have gained much interest in computer vision and machine

learning [Wan et al., 2014, Karpathy and Fei-Fei, 2015, Gordo et al., 2016]. These methods have shown increased effectiveness and efficiency in classification, recognition and retrieval tasks. Unlike conventional machine learning methods many deep learning methods attempt to model high-level abstractions in data.

2.5 An Overview of the Content-Based Image Retrieval Process

We have studied existing CBIR systems and now refer to the process of CBIR. Four main components of a CBIR were mentioned in section 2.2 and that can be further subdivided *Image Database*: as data collection, feature extraction, building up feature database, *Information Need*, *Result Generation and Presentation*: as searching and arranging the order for presentation to the user, and *Relevance Feedback* for ease of understanding.

2.5.1 Data Collection

Data collection can be conducted by using a web crawler program to collect images from the internet. Already there are standard datasets available, such as the COREL dataset [Tao et al., 2007], the Caltech-101 [Fei-Fei et al., 2004] and Caltech-256 [Griffin et al., 2007] datasets, the Oliva and Torralba [Oliva and Torralba, 2001] dataset, the Wang [Wang et al., 2001, Li and Wang, 2003] dataset, the Nister [Nister and Stewenius, 2006] dataset, the Flickr [Huiskes and Lew, 2008] dataset, the SUN [Xiao et al., 2010] database and the Tiny Images dataset [Torralba et al., 2008a], consisting of some 80 million images. These are the datasets mostly used in literature.

2.5.2 Feature Extraction, Selection and Representation

CBIR consist of several steps and each step needs to be considered when developing a CBIR system. Firstly image descriptors must be selected according to the database that is going to be addressed. All the colour, texture and shape features are useful when describing image content in general image datasets, as general datasets may be comprised of colour images, images with texture and images with objects. However, all the selected features may not be useful due to noise and redundancy, which leads to false matches and hurts the retrieval performance

when they are used together. Therefore, feature selection is necessary to eliminate useless features and select the best combination of features to improve retrieval performance. These selected features must be represented in a certain way in order to be used in the CBIR process and this is mainly divided into two categories ,namely global representation and local representation based on the manner of feature extraction.

Features and Descriptors

Images are generally represented as a collection of feature vectors in the database and are retrieved according to relevance by calculating similarity measures between the query image features and the target image features in the database. It is, therefore, necessary to build the feature set or image signatures; image signature is any representation which is generated by extracted features in order to use in search process. Firstly, system features must be selected from various types of available features. It is common to use low-level features like colour, texture and middle-level features such as shape in current CBIR systems. Some systems use only one feature for retrieval- colour [Javidi et al., 2008,Murthy et al., 2010,Sharma et al., 2011], texture [Kokare et al., 2005], structure, edge. While most systems use a combination of features like colour and texture [Carson et al., 2002,Fanhui, 2011, Jiang et al., 2006,Liu et al., 2008,Gebara and Alhajj, 2007,Heller and Ghahramani, 2006], colour and shape or texture and shape [Iqbal and Aggarwal, 2000,Takahashi et al., 2000]. To achieve a better performance, it is better to use a combination of features [Hiremath and Pujari, 2007b, Wang et al., 2001, Kherfi and Ziou, 2006, Iqbal and Aggarwal, 2003, Abubacker and Indumathi, 2010, Tahoun et al., 2005] rather than only one or two. As an example, colour histograms have been used in many applications, but a colour histogram alone cannot describe the image well. It does not describe the distribution of the colours and there may be totally different images with the same colour histogram.

When deriving features, feature descriptors play a major role in extraction. Feature quality can vary according to the selected descriptors. Good visual descriptors should be invariant in scale changes, translation and rotation, as well as different lighting conditions and viewpoint changes. There are different kinds of descriptors available to describe features.

Colour Descriptors

Colour is the most significant feature in an image and it is one of the most simple and extensively-used features in image retrieval [Carson et al., 2002, Fanhui, 2011, Javidi et al., 2008, Murthy et al., 2010, Jiang et al., 2006, Liu et al., 2008, Gebara and Alhajj, 2007, Sharma et al., 2011, Heller and Ghahramani, 2006]. It is relatively robust to background complication and independent of size and orientation of the image. Before extracting a colour feature, it is necessary to select a suitable colour space and representation method (descriptor). There are several colour spaces but RGB, CMY, YCbCr, HSV, Lab and LUV are the most significant among them. After selecting the space, colours are defined from that. An appropriate colour descriptor can then be selected from the colour histogram (conventional colour histogram, fuzzy colour histogram, cumulative colour histogram), colour correlogram, colour moment, colour coherence vector and colour palettes [Hiremath and Pujari, 2007b].

Texture Descriptors

Texture is another fundamental and interesting feature which has been used in image retrieval literature. Although texture is not well defined like colour, it does describe the content of many real-world images and provides important characteristics for surface and object identification. Texture refers to repetitive patterns employed in images, such as clouds, trees, fabrics, sand or bricks. Texture representation is more important in the fields of texture classification, image segmentation and image shape identification. A large variety of texture descriptors, such as wavelet Transform, Gabor wavelet, Edge Histogram Descriptor, co-occurance Matrix, Tamura features have been developed, as they play an important role in computer vision. These methods can be subdivided into three categories - statistical methods, structural methods, and spectral methods [Takahashi et al., 2000]. The evaluations of these descriptors can be found in [Pichler et al., 1996, Howarth and Ruger, 2004].

Shape Descriptors

The shape feature is fairly well defined but it is shown to be very useful

in manmade objects, especially when a query image is drawn by hand and if there is only one object in the image. Image features are invariant to translation, rotation and scaling if they are extracted by segmented regions. However, segmentation is not precise and there are still no methods to segment images accurately. There are many shape descriptors, and the evaluations of those descriptors can be found in [Kidiyo and Joseph, 2008, Amanatiadis et al., 2011, Amanatiadis et al., 2009]. Shape representation can be mainly divided into two categories - boundary-based and region-based methods. Boundary-based methods are based on the boundary of the shape by neglecting what is inside. The region-based methods are based on the entire region of the shape.

Other Descriptors

Global - Gist: GIST is a global descriptor proposed for the recognition of real-world scenes by Oliva and Torralba [Oliva and Torralba, 2001] and used in scene recognition applications [Torralba et al., 2008a, Li et al., 2008b, Hays and Efros, 2008] and have shown good retrieval performance. Its idea is to generate a very low dimensional representation of the scene which does not require any form of segmentation. They proposed a set of perceptual dimensions (naturalness, openness, roughness, expansion and ruggedness) to represent the dominant spatial structure of a scene and these dimensions can be reliably estimated using spectral and coarsely-localised information [Oliva and Torralba, 2001]. They named it Spatial Envelope.

Feature Selection

There are a number of features that can be extracted from images using different descriptors. As we cannot use all the features in a CBIR system, the relevant features must be selected. Therefore, feature selection is an important step in CBIR and the best descriptive features are selected among many features. It reduces the number of features, removes irrelevant, redundant or noisy data, and brings the immediate effects for applications. This can be done offline as well as online. The objective of feature selection includes improving retrieval performance in terms of accuracy and speed by removing less relevant or noisy features. Feature selection methods are often divided into three categories: filter methods, wrapper

methods and embedded methods [Liu et al., 2013, Zhao et al., 2013b, Farahat et al., 2013]. General feature selection steps can be seen in figure 2.3.

Filter methods [Yu and Liu, 2003, Peng et al., 2005, Estevez et al., 2009] select the features in a manner that is independent of the classifier. This does not take into account feature interactions and is generally not a recommended way of doing feature selection as it can lead to lost information. Common measures used in filter methods include mutual information and a correlation coefficient.

Wrapper methods [Kohavi and John, 1997, Maldonado and Weber, 2009, Wang et al., 2014] use the learning machine of interest, as a black box to score subsets of features according to their prediction performance. An example would be how Random Forest [Gharsalli et al., 2015, Murata et al., 2015] is widely used by the competitive data science community to determine the importance of features by looking at information gain and leave-one-out [Liu et al., 2013]. In this method, computation time is high when the number of variables is large but it usually provides the best performing feature set for that particular type of model.

Embedded methods [Das, 2001, Guyon and Elisseeff, 2003, Roth, 2004, Yuan et al., 2013a] involve carrying out feature selection and model tuning at the same time. They try to combine the advantages of both previous methods. Some methods include greedy algorithms like forward and backward selection as well as Lasso [Roth, 2004, Yuan et al., 2013a] and Elastic Net [Zou and Hastie, 2005] based models.

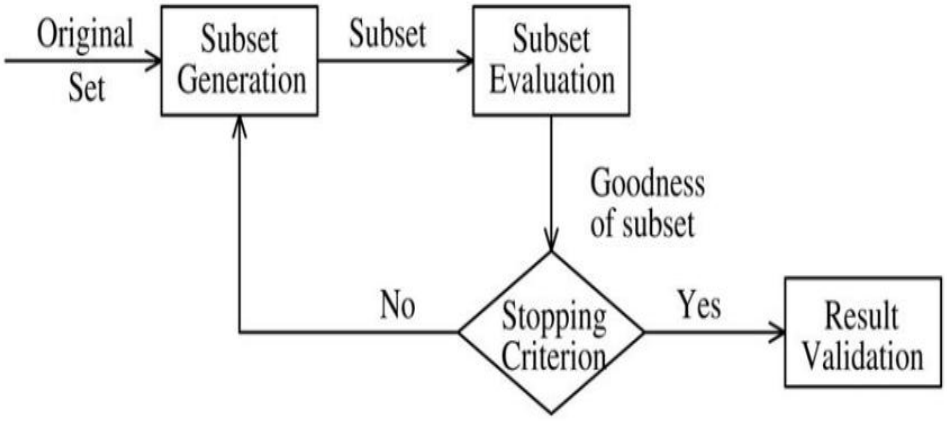


Figure 2.3: Steps of feature selection [Liu and Yu, 2005].

Image Representation

There must be a way to represent selected features in CBIR. In literature, many techniques have been used to represent the image content. It can be mainly divided into two approaches - local and global. Feature extraction can be done both locally and globally. Local features are the low-level features which are extracted from sub-images, segmented regions or points of interest, while global features are the same type of features which are extracted from the whole image without subdividing, as represented in figure 2.4.

Beside this, there is a precise way to represent images called Bag-of-Features (BoF), also known as Bag-of-Words (BoW), which quantises feature descriptors to generate words and represent images by histogram of those words of which we will give a detailed overview in Section 2.5.3.

Figure 2.4.a shows a global representation of the image using low-level features. Here, feature extraction is done for the full image. Figure 2.4.b shows a local representation of the image using low-level features. Here, feature extraction is done for each image block separately and finally combines them together.

Global Representation

In global feature presentation, image content is described by low-level features such as colour, texture and shape. The global feature approach achieves a good performance in some cases but not for all image and query types. Moreover, global features do not work for databases with many image categories and cannot distinguish the foreground from the background of an image and mix information.

Local Representation

Local features are very useful in extracting the tiny details of the image and withstanding translations of images. Local representation has been applied to a wide range of CBIR systems and applications to achieve robust representation.

The key proprieties of a good local feature are [Estrada et al., 2004]:

- Must be highly distinctive - a good feature should allow for correct object identification with a low probability of mismatch.

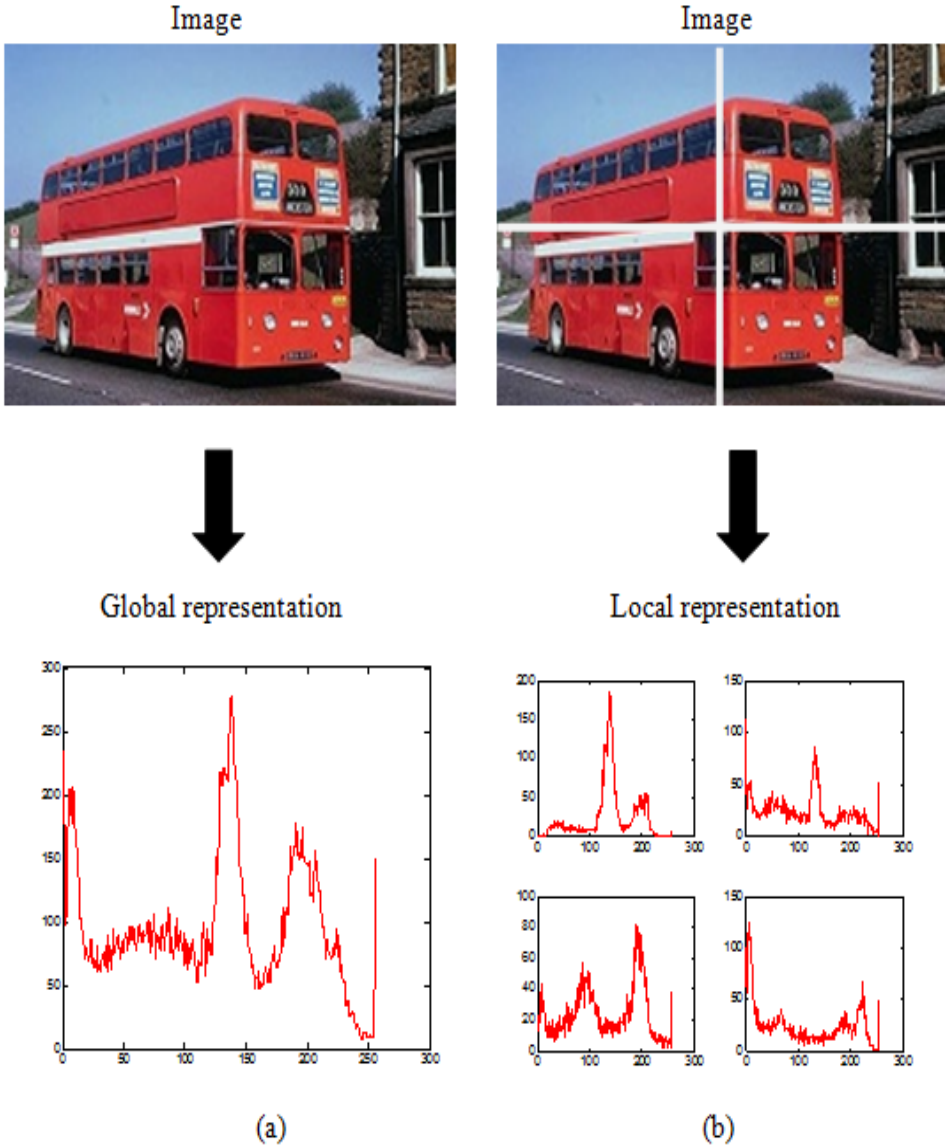


Figure 2.4: An example of the global and local representation of images using a colour histogram. The x-axis is an index into a colourmap.

- Should be easy to extract.
- Invariance - a feature should be tolerant to image noise, changes in illumination, uniform scaling, rotation and minor changes in viewing direction.
- Should be easy to match against a large database of local features.

Local representation has two types of approaches - dense and sparse. Sparse representation decomposes the image into localised patch descriptors around interest points (key points), while dense representation divides the image into localised patch descriptors on a regular grid [Jain, 2010].

2.5.3 Bag of Features Approach (BoF)

During the past decade, the popularity of the BoF approach in fields of classification and retrieval in CBIR [Yuan et al., 2011b, Takahashi et al., 2000, Aman et al., 2010, Zepeda et al., 2009, Zheng et al., 2006, Kogler and Lux, 2010, O'hara and Draper, 2010] is immense because of its simplicity and performance. This approach was introduced by Sivic and Zisserman [Sivic and Zisserman, 2003] to the computer vision community and was inspired by the BoW model in text document retrieval. The BoF approach is analogous to the BoW representation in text document retrieval. The image features represent the local areas of the images, just as words represent the local features of a document. The visual vocabulary or visual-codebook is raised by clustering image features that are extracted from images in the database and it is analogous to the vocabulary in text retrieval, which is derived from the corpus. Firstly, features are clustered and each cluster stands for a visual word. The term vector that represents the image is a sparse vector with all assigned codebook IDs or visual words and it is analogous to the term vector that represents a document in text retrieval. Codebook size (vocabulary size) is predefined by corpus in text retrieval, while in CBIR, it depends on the application, and it may vary from a few hundred to millions. This process of BoF representation is shown in figure 2.5.

BoF approaches do not take into account the spatial information, relative location, scale and the orientation of features. It is a collection of image features without having any order (Bag), nevertheless, it is a powerful representation that has shown better performance for image classification and image retrieval than other approaches, and has become well-established. The BoF image representation includes following steps:

1) Local patches detection.

2) Compute local descriptors on them.

3) Quantisation of the descriptors to obtain words in order to create a visual

vocabulary.

4) Assign terms to regions in an image by a similarity measure.

5) Record the occurrences of each word that appears in the image.

One of the key points is to select the appropriate feature descriptor for the BoF representation. The most popular feature descriptor in the BoF approach is the SIFT descriptor because it often outperforms other descriptors [Sivic and Zisserman, 2003, Aman et al., 2010, Zepeda et al., 2009, Zheng et al., 2006]. However, other descriptors also have been used in the BoF representation.

Vector quantisation (clustering) is widely used in an image retrieval task, especially when using a large set of data. Clustering is used to group image features as well as datasets according to the distribution of features. It is used to build the visual vocabulary in BoF algorithms. It clusters the feature descriptors of all small sub-images and each cluster represents a visual word. There are many clustering algorithms. K-means is the one which is widely used [De Vries and Geva, 2009a, Yang et al., 2009, Murthy et al., 2010] for clustering because of its accuracy and usability. It was developed by J,MacQueen [MacQueen, 1967]. The output of k-means has a great effect on initial data centres and this is not suitable for very large databases. There are improved different versions of k-means (hierarchical k-means- [Murthy et al., 2010, Nister and Stewenius, 2006], fuzzy k-means [Li et al., 2008a, Wang et al., 2008]) introduced for clustering in CBIR.

Even though K-means has better accuracy, it is not suitable for very large datasets. K-tree is a height-balance cluster tree introduced by Geva [Geva, 2000]. It is a combination of B-tree and k-means algorithms. It is mostly suitable for large collections due to its low complexity and it supports online dynamic tree construction and multi-granularity clustering. It has been used and achieved good performance in text retrieval [De Vries and Geva, 2009a, De Vries and Geva, 2009b, Geva and De Vries, 2011, Vries et al., 2009]. According to our knowledge, it has not been used in image retrieval applications.

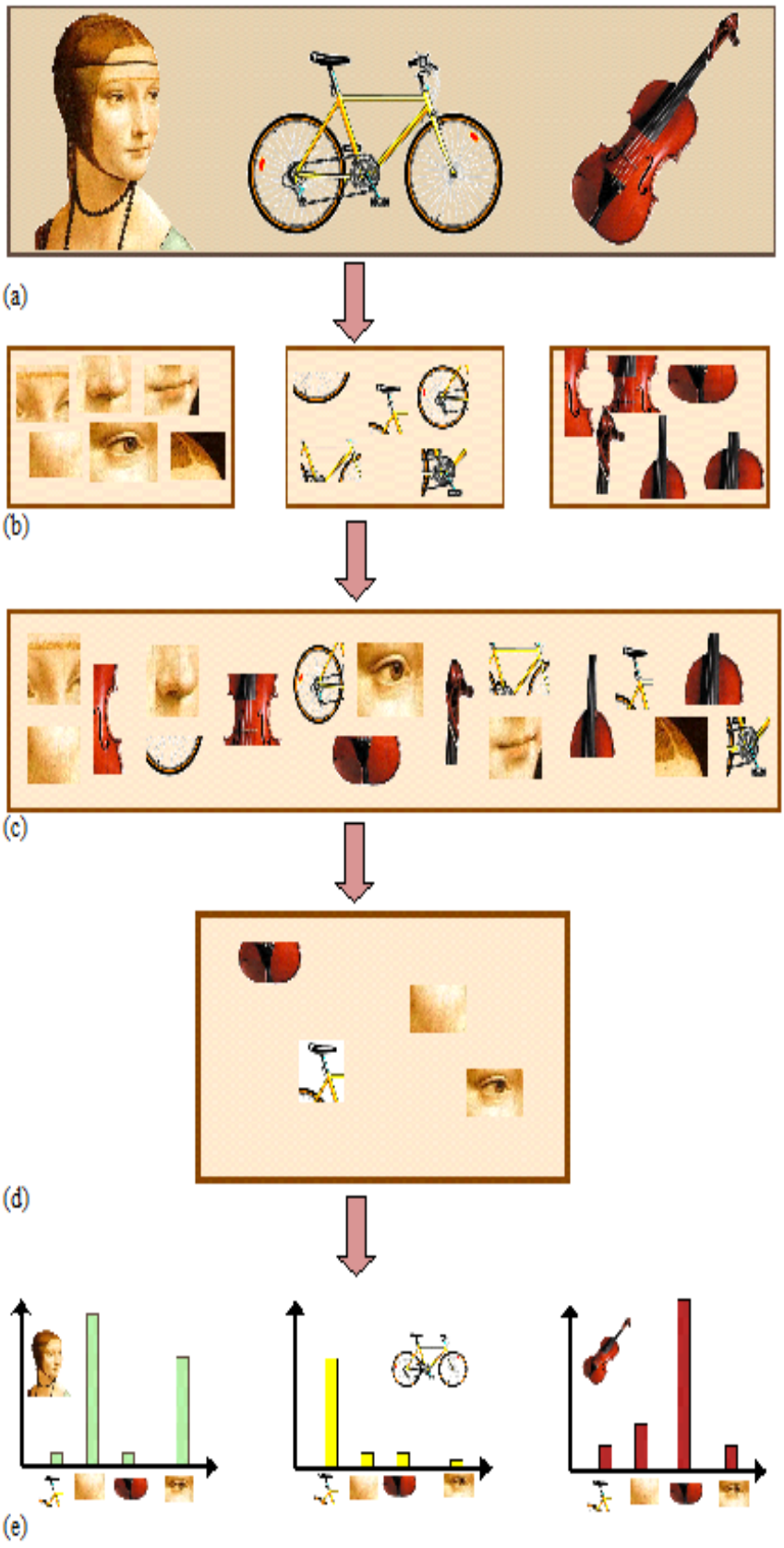


Figure 2.5: An example of a BoF model. (a) Database of images (b) Feature extraction of each image (c) Collection of features for a whole dataset. (d) Generate codebook or vocabulary and represent each cluster. (e) Represent each image using a generated codebook.

According to the [Geva, 2000], the K-tree algorithm (order m) can be defined as follows:

- All leaves are on the same level.
- All internal nodes, including the root, have at most m nonempty children, and at least 1 nonempty child.
- Codebook vectors (clusters) act as search keys.
- The number of keys in each internal node is equal to the number of its nonempty children, and these keys partition the keys in the children to form a Nearest Neighbour (NN) search tree.
- The level immediately above the leaf level forms the clustering codebook level.
- Leaf nodes contain data vectors, or references to data vectors.

K-tree uses Euclidean distance for all measures of similarity and the search path is determined by NN search.

When generating vocabulary, there are two types of weighting schemes in CBIR as a soft assignment (Fuzzy codebook [Zepeda et al., 2009, Kogler and Lux, 2010]) and a hard assignment to visual words (terms) [O'hara and Draper, 2010]. It is found that the soft assignment approach improves retrieval accuracy more than the hard assignment [Philbin et al., 2008] and some publications concluded that "the soft assignment of key points to visual words is superior to hard assignment" [Kogler and Lux, 2010].

CBIR systems generate visual vocabularies using these vector quantisation methods. As mentioned earlier, the BoF approach is order-less and Spatial Pyramid Matching was introduced to improve the efficiency of the BoF approach by adding spatial data which was missing in the BoF by Lazebik [Lazebnik et al., 2006, Lazebnik et al., 2009]. A histogram of visual words is computed for each image sub-region at each resolution according to that and figure 2.6 shows the spatial pyramid representation for two levels.

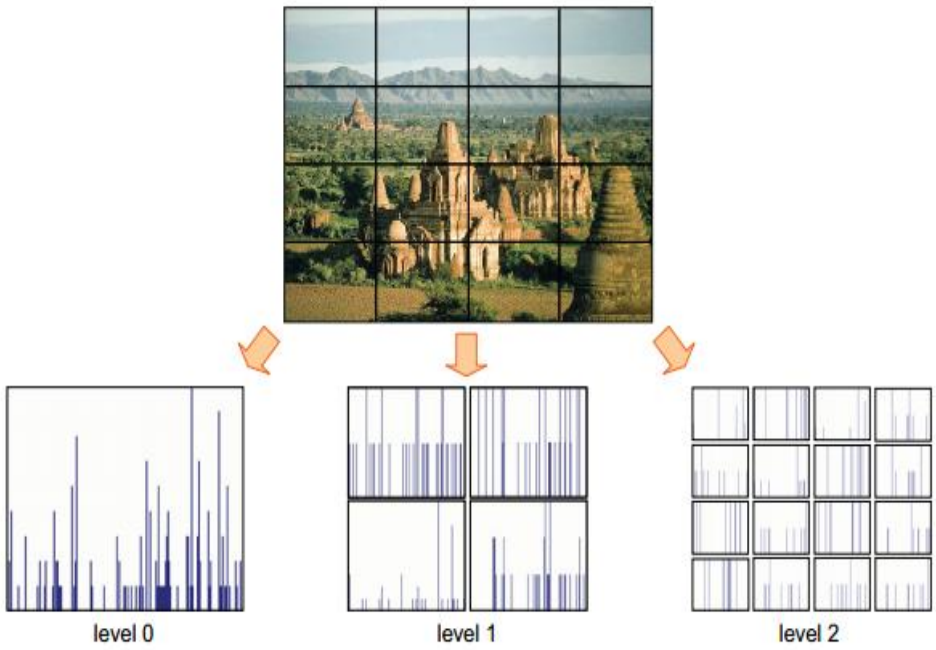


Figure 2.6: A schematic illustration of the spatial pyramid representation. A spatial pyramid is a collection of order-less feature histograms computed over cells defined by multilevel recursive image decomposition. At level 0, the decomposition consists of just a single cell, and the representation is equivalent to a standard bag of features. At level 1, the image is subdivided into four quadrants, yielding four feature histograms, and so on [Lazebnik et al., 2009].

Pyramid matching works by placing a sequence of increasingly coarser grids over the feature space and taking a weighted sum of the number of matches that occur at each level of resolution. At any fixed resolution, two points are matched if they fall into the same cell of the grid. Matches found at finer resolutions are weighted more highly than matches found at coarser resolutions because higher levels provide a more precise representation than lower levels [Lazebnik et al., 2009].

Figure 2.7 gives an example of a construction pyramid up to level 2. It is found that for strong features, image subdivision till level 2 is enough and there is no performance improvement in level 3 because it is too finely subdivided and it yields too few matches [Lazebnik et al., 2006]. Finally, the feature vector is generated by combining all the histograms in different levels and it is named the Pyramid

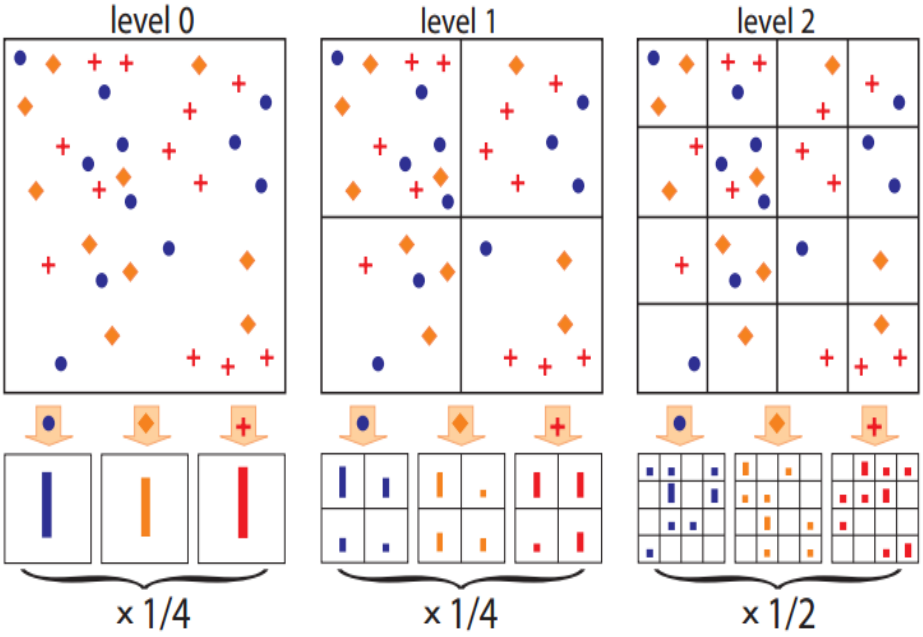


Figure 2.7: An example of constructing a pyramid for L = 2. The image has three feature types, indicated by circles, diamonds, and crosses. At the top, the image is subdivided at three different levels of resolution. Next, the features that falls in each spatial bin for each level of resolution and each channel are count. Finally, each spatial histogram is weighted [Lazebnik et al., 2006].

Histogram Of visual Words (PHOW).

Sivic and Zisserman [Sivic and Zisserman, 2003] showed that the most frequent visual words occurring at many places in the image are responsible for mismatches, so simply eliminate any too-frequent terms that appear in many images [Sivic and Zisserman, 2003]. Moreover, it reduces the size of the vocabulary. To achieve better quality, term weighting is applied in BoF. Weighting helps to improve precision and recall. This is the motivation behind Term Frequency-Inverse Document Frequency (TF-IDF) and log-likelihood [Chappell, 2015] techniques used in text retrieval. Term weights can be computed in many different ways like Term Frequency (TF), TF-IDF, binary weighting [Sivic and Zisserman, 2003, Yang et al., 2007], and log-likelihood [Chappell, 2015] . Log-likelihood [Chappell, 2015] has shown better performance in text retrieval and it has not been applied to the BoF/BoW in image retrieval.

A distance measure is also needed in term vector space for measuring similarity between two images for classification and retrieval. Euclidean (L2) [Laaksonen et al., 2002, Saha et al., 2007] and Manhattan (L1) distances are popular choices because these distance measures and distances can be calculated efficiently on sparse term vectors using inverted indexes.

2.5.4 Build-Up Database and Indexing

Most of the CBIR systems use a certain number of aforementioned feature descriptors to extract the image content. All the features are extracted from all the images in the dataset and stored as feature vectors to represent each image by a vector. However, such feature vectors tend to be very high dimensional, which may lead the system to suffer from "curse of dimensionality". This leads to a decline in the efficiency of CBIR and moreover, makes efficient CBIR unfeasible for very large image collections due to the high computational time. Furthermore, the memory requirement for the storages increases. This is n vital issue which has been commonly tackled by embedding a dimension reduction mechanism on feature vectors before setting up an efficient indexing scheme.

Dimensionality Reduction

Traditionally, CBIR systems are characterised by slow response times or high computational costs due to the high-dimensionality of the feature spaces used to describe images. High dimensional indexing is one of the prevailing challenging tasks in CBIR. This problem has been tackled by a number of CBIR techniques for dimensionality reduction, including latent semantic analysis (LSA) [Gorman and Curran, 2006], Principal Component Analysis (PCA) [Banda et al., 2013], Singular Value Decomposition (SVD) [Banda et al., 2013] and Locality Sensitive Hashing (LSH) [Gorman and Curran, 2006]. All these techniques share the same objective of reducing the dimensionality of the feature vectors while maintaining as much as possible the fidelity of the information included in the descriptors.

Random Indexing (RI) is a technique for dimensionality reduction that has been widely used in text retrieval but that has received lesser attention in CBIR [Vries et al., 2009, Geva and De Vries, 2011]. Random indexing relies on random projections to avoid performing computationally expensive matrix factorisations that are

instead performed by techniques like SVD, LSA and PCA. Compared with other methods, RI has a lower computational cost and a lower complexity, but it also has a competitive accuracy, and, most importantly, it is an incremental approach, which allows the index streaming of objects as they are included in the collections without requiring the re-indexing of older objects [Vries et al., 2009, Sahlgren, 2005].

In dimensionality reduction, image signatures can be generated in binary format using hash functions namely Binary Image Signatures. Hash functions generate a fixed-length binary string from the given input. Robust hash functions in retrieval are able to produce identical signatures for two similar inputs in a lower-dimensional space which are non-invertible. On the other hand, different inputs will produce vastly different output signatures. In CBIR, binary signatures can be used to compare similarity in content even though the content is not reproducible from the signature. Searching is much faster than using the original representation, as signatures are much smaller than the original full-dimensional vector. Robust image hashing was introduced in late 90's and has been used in applications other than CBIR, such as image watermarking and biometric template security, while the purposes are different from CBIR. In CBIR, binary signatures are used to acquire higher efficiency [Torralba et al., 2008b, Lv and Wang, 2013, Liu et al., 2014, Yang et al., 2015b, Khavare and Manjrekar, 2015, Feng et al., 2016] and the latter methods use hashed for authentication/cryptography [Li and Wang, 2016, Tang et al., 2016, Zhao et al., 2013a, Wang et al., 2015, Yang et al., 2015b, Khelifi and Jiang, 2010, Chen and Chandran, 2010]. In authentication, hash functions use an 'avalanche effect', which is sensitive to even the slightest changes in the input, and means that a very small change in input (nearly similar) will produce a large change in the binary signature. This feature is quite important for cryptography. However, this is not the target of CBIR, and CBIR needs to use hash functions to generate similar signatures for similar inputs, which means that slightly different inputs create only slightly different signatures (nearly similar) as its task is to find similar matches. LSH was introduced to reduce the dimensionality of high-dimensional data by producing binary signatures (short hashes) to search accurately and efficiently, which is useful in CBIR. LSH differs from the conventional and cryptographic hash functions mentioned above.

Reducing the dimensionality of features will have considerable effect, not only on similarity measures but also when storing. So it can achieve better compression effectiveness, as the searching time and memory requirement reduces with a lower-dimensional vector, rather than a full-dimensional vector.

Image Features Storage - Indexing

Early image retrieval systems used simple files in a directory or entries in databases to store the extracted visual descriptors of an image such as mySQL and DB2 [Flickner et al., 1995]. However, they performed very poorly from a computational perspective as most of these systems only used linear searches and used long real-valued vectors.

Later tree-based mechanisms were introduced for indexing, such as R-tree [Guttman, 1984] and its different versions, and k-d trees [Egas et al., 1999]. An overview of several tree structures and their properties can be found in [Smeulders et al., 2000] for more detail. Inverted files that have proven to be very useful for text retrieval [Moffat and Zobel, 1996] are shown to be efficient also for image retrieval when the feature space is only very sparsely populated [Squire et al., 1999, Zhang et al., 2009]. For binary image representation, a still sequential search works well, as not much computational time is necessary. Recently, an Inverted Signature SLicing (ISSL) mechanism was introduced for indexing binary image signatures when the datasets become sufficiently large [Chappell et al., 2013]. ISSL searches signature files in a fraction of the time, but there is a trade-off between the speed and the quality of the retrieval.

2.5.5 Information Need

CBIR is popular in different areas and images are generally required for different reasons. There are different systems generated for field specific tasks as well. Some of the areas in which CBIR is widely used are medicine, home entertainment, web-searching, journalism, television and advertising, economy, sports and security. Even in one field, requirements may be different . For example, in home entertainment, sometimes the user may need to find a particular event, a particular scene, specific person, and so on.

If the user needs to find it manually, it is really hard and time-consuming with

the load of data, and if searching is done by text related to images, the user must make sure he/she tags and provides a description for each image every time. It is efficient and effective if the user can search images by providing a sample image which represents the images which he/she requires. If the system retrieves images by using a query as an image it is called query by example paradigm. This query image may be an image from the database or an image from anywhere. Some systems use one query image, while other systems use several images as a query. This may be a sketch in some cases.

This is how the user tells the system about his/her information need. Nowadays, most systems depend on QBE without depending on the text.

2.5.6 Result Generation

Once the information need is given, the retrieval results must be displayed. In CBIR systems, a search engine retrieves the list of images similar to the query image, ranked according to a similarity measure. Similarity measuring can be done in various ways according to the image representation and descriptors used in the system. One of the major aspects that affect the precision of the retrieval process is the similarity measure. To achieve better precision, a similarity measure must be selected by considering image descriptors and image representations. Different systems use different distance measures, such as Euclidean distance [Laaksonen et al., 2002], Weighted Euclidean distance, Manhattan distance, Cosine similarity, Hausdorff [Veltkamp, 2001], Mallows distance, Kullback-Leiibler (K-L) divergence [Long et al., 2003], Jaccard co-efficient, Minkowski form distance [Veltkamp, 2001,Carson et al., 2002], Mahalanobis distance [Carson et al., 1997] and histogram intersection [Saha et al., 2007]. Hamming distance is used when images are represented in binary form.

In terms of querying speed, users always prefer faster systems. The simplest approach to search NN is the linear search, which increases computation time with image database size, so it is prohibitive for large databases. Thus, to give faster responses to the user, there must be a way to avoid scanning the entire database every time a query is submitted. Therefore, clustering is applied to the final representation, so that similar image signatures (full-feature vectors) are gathered around one point. When querying, the similarity between the query and

the cluster centres are found and the similarity between the query and each image in the selected cluster is compared, thereby narrowing the search. Finally, the list is ranked according to the similarity measure. Most systems use Euclidean distance to measure the similarity. There are some algorithms which follow the above sequence and overcome problems in NN like KD-tree, and Approximate Nearest Neighbor (ANN) [Ferhatosmanoglu et al., 2001]. Some systems use binary code by mapping feature space into binary values which helps to improve efficiency as well as memory usage [Torralba et al., 2008b, Peker, 2011, Zhen-Sheng, 2012, Lv and Wang, 2013, Liu et al., 2014, Yang et al., 2015b, Khavare and Manjrekar, 2015, Feng et al., 2016].

Geva and De Vries [Geva and De Vries, 2011] introduced TopSig in text retrieval literature, which is a topological preserving document signature. It has been shown that it can achieve performance in indexing and retrieval with a document signature similar to inverted file indexing. Once the image signatures are clustered, it can be used for searching. Unlike inverted file representation it has the advantage that it is not a distributed representation - a signature is associated with an image and it is possible to associate multiple signatures with an image. The system was shown to be highly effective and efficient in applying focused user RF in text documents by highlighting passages of relevant text in a document. This has achieved good performances in text retrieval and has not applied to image retrieval applications yet.

Once a system has ranked the images in a database using a similarity measure for a given query, the results are then displayed to the user. Most retrieval systems display the images separated over several pages of result. The number of images shown on a page varies, however, most CBIR systems show 20 images in a grid-based presentation. Visualisation is also important in a user's point of view as it is convenient for the user.

2.6 Relevance Feedback (RF)

The semantic gap is the gap between the high-level semantic concepts required by the system users and the low-level visual features extracted from images. It is used by the retrieval algorithm to find the similarity between images. Semantic gap is the basic and openly challenging problem in CBIR and has attracted a lot

of attention from the research community. Many researchers have addressed this problem from different points of view. Different techniques have been applied in CBIR systems [Liu et al., 2007] to achieve high level semantics. Some methods that have been used to reduce the semantic gap are the exploration of domain knowledge to define object ontology for image annotation using machine learning techniques such as Support Vector Machine (SVM), decision trees, k-means, and others to associate low-level features with high-level semantics [Laaksonen et al., 2002, Liu et al., 2008], introducing RF into the retrieval loop for a continuous learning of the user's intention [Rui et al., 1997, Su et al., 2003, Jiang et al., 2006, Saha et al., 2007, Kherfi and Ziou, 2006, Fanhui, 2011, Cui and Zhang, 2007, Jing et al., 2002, Wang et al., 2006, Ko and Byun, 2002], making use of both textual information obtained from the web and the visual content of images [Cai et al., 2004] and generating Semantic Templates (ST). Among these methods, RF is an online processing algorithm which interacts with users while others are engaging in offline processing.

RF [Harman, 1992] is an approach that seeks to improve the precision of search results through the incorporation of user feedback in response to an initial result list. This was originally developed for text-based Information Retrieval (IR) systems to improve the quality of retrieval. It was introduced to CBIR as a means of improving the effectiveness of the system by bringing the user into the retrieval loop and it has shown a significant boost in the performance of CBIR systems [Cui and Zhang, 2007, Jing et al., 2002, Wang et al., 2006, Ko and Byun, 2002, Su et al., 2011a, yi Lee and Lee, 2013]. When interacting with an RF interface, the user informs the search engine whether the results that have already been provided are useful or not.

In a typical RF process, in any given image query, the system provides an initial list of ranked images according to some similarity measure [Harman, 1992]. From the retrieved images, users are required to select relevant images as positive feedback and irrelevant images as negative feedback (binary RF), or alternatively, rate each image, for instance, from one to five (1-irrelevant, 5-most relevant)(weighted RF). The system exploits this feedback to refine the original query and retrieve a new list of images, which are intended to be more relevant than those displayed before the RF interaction, because they would have been retrieved using a bet-

ter representation of the user's information needs and preferences towards certain types of images. This process may be repeated until the user is satisfied with the retrieval results or issues a new query.

However, this RF can be divided mainly into two - Pseudo relevance feedback (PRF) and interactive RF which will be discussed in the following sections.

2.6.1 Pseudo Relevance Feedback (PRF)

Despite the volume of work previously done with RF [Cui and Zhang, 2007, Jing et al., 2002, Wang et al., 2006, Ko and Byun, 2002] (user-interactive) in CBIR, it still a challenging task to develop an efficient and effective RF mechanism. Intuitively, the RF mechanism places a burden on users as they need to provide explicit feedback, while they may be reluctant to do so, and providing feedback may be cognitively demanding if many images need to be rated according to relevance, or when images are not clearly relevant or irrelevant. Indeed, recent work has grounded this intuition into a formal model of information-seeking and retrieval based on the Economics model of interactive IR [Azzopardi, 2011, Azzopardi and Zuccon, 2015] and showed what gains RF needs to achieve in order to be "economically" (Where cost and benefit are modeled in terms of interaction time) acceptable to the user. Furthermore, the RF process typically occurs over several iterations and users may not be willing to put up with this. On the other hand, the RF mechanism also puts burdens on the search system. In fact, the search process requires more run-time when RF is used, because of the usually larger (less sparse) representation of the query. In addition, the system may not be able to provide satisfactory results in the presence of only limited interactions, e.g. when the user selects only one or two images in the RF process.

PRF approaches have been introduced with the aim of removing the need for the users to explicitly select relevant documents for RF, but yet retain some of the effective improvements the feedback mechanism delivers without the need for user interaction. This is achieved automatically by considering the first N (top-ranked) retrieved documents in the first pass of retrieval using the original user's query as relevant, and use this information to produce a refined search whose results are then shown to the user. PRF has been widely used in text retrieval [Huang et al., 2006, Liua et al., 2007], image retrieval [He et al., 2009, Lin et al., 2003, Yap and

Wu, 2007] and video retrieval [Rudinac et al., 2009, Yan et al., 2003]. In addition, a number of PRF techniques also consider negative feedback [yi Lee and Lee, 2013], which is implicitly acquired using the last N (bottom-ranked) images from the initial result list, which are therefore assumed to be irrelevant. PRF is a black-box to the users. Most PRF techniques assign the same importance to every document when reformulating the query through the RF mechanism.

2.6.2 Interactive Relevance Feedback

The Rocchio algorithm [Rocchio, 1971] is a widely-known RF mechanism that is commonly applied independently of the document media (i.e. text, images, videos, etc.). In the Rocchio algorithm, documents and queries are represented as vectors of features, following the vector space model for IR [Rocchio, 1971, Ishikawa et al., 1998, Lu et al., 2000, Shaw, 1995]. The intuition behind the algorithm is that feedback about relevant documents is used to move the vector of the query closer to the vectors representing the documents that have been marked relevant; vice-versa, feedback about irrelevant documents is used to move the query vector apart from the vectors of those irrelevant documents. Typical document retrieval systems integrate a measure of RF into the vector space model by Rocchio's formula [Rocchio, 1971] defined as in equation 2.1. Formally, this is summarised in the equation:

$$q_m = \alpha q_0 + \beta \left(1/|D_r| \sum_{d_j \epsilon D_r} d_j \right) - \gamma \left(1/|D_{nr}| \sum_{d_j \epsilon D_{nr}} d_j \right)$$

where q_0 is the original query vector and q_m is the revised query representation after RF, D_r and D_{nr} are the sets of known relevant and irrelevant documents, respectively, and α, β and γ are balancing weights. The Rocchio algorithm is at the basis of many CBIR systems, e.g. [Ishikawa et al., 1998, Lu et al., 2000]. Empirical results have shown that this RF techniques lead to improvements in retrieval effectiveness.

Other RF methods have considered feedback as a two-class classification problem, where positive feedback is grouped into one class and negative feedback into another. This intuition allows for the application of binary classifications methods, like biased subspace learning, to the RF task. Methods in this family include

Biased Discriminant Analysis (BDA) [Zhou and Huang, 2001a], Kernel Biased Discriminant Analysis (KBDA) [Zhou and Huang, 2001b], Direct Linear Discriminant Analysis (DLDA) [Yu and Yang, 2001], Direct Kernel BDA (DKBDA) [Tao et al., 2006], and Marginal Biased Analysis (MBA) [Xu et al., 2007]. These latter examples further divide the negative class into a number of sets to avoid the problem of having too many negative examples over a positive one, which would bias the classifiers and let negative examples dominate. These methods, however, do not allow for the specifications of different preferences or strengths with respect to the positive RF (i.e. different ratings for relevant images) and thus all relevant (positive) images are treated equally. A variety of RF systems have been designed to bridge the semantic gap between low-level visual features and high-level semantic concepts for an image retrieval task. However, it must be noted that these proposed algorithms are commonly evaluated using PRF settings or simulated interactions and neglect experimental evaluation with real users. A more comprehensive overview of feedback methods can be found in [Li and Allinson, 2013] as Lu and Allinson provide a detailed survey of RF mechanisms for CBIR.

As a summary, existing RF systems work with real value vectors [Wu and Yap, 2006, yi Lee and Lee, 2013, Yap and Wu, 2007, Lin et al., 2003, Yan et al., 2003]. Some of the existing systems [Wu and Yap, 2006, yi Lee and Lee, 2013, Yap and Wu, 2007] use SVM which need to be trained and are computationally expensive. Existing PRF systems consider every document with equal importance when generating new queries from RF. This assigns undue importance to less likely-relevant results, particularly when the PRF result list is long. Moreover, the computational time of these RF mechanisms is high, as every time, it has to go back to the original feature space and search the full database in each iteration.

2.7 Chapter Summary and Conclusions

This literature review focused on current CBIR systems and the existing techniques used to reduce the semantic gap. This chapter explored the main concepts in CBIR. Initially, a general introduction was given to the concepts and an overview of the challenges was explained. Then a number of well-known and most-cited image retrieval systems were presented with their system features. Existing technologies and gaps in each step in CBIR were discussed specifically as data collection,

feature extraction, selection and representation, indexing, information need and results generation. A comprehensive overview of the image descriptors that can be extracted from images and their selection methods to select the best subsets for CBIR systems, and representation methods were discussed in the above sections. The BoW/BoF approach was discussed specifically. Moreover, information on how to address the curse of dimensionality and indexing was provided. How user information need is taken in the CBIR, major similarity measures, and results representation were also covered in those sections. As CBIR systems are generated for real users, RF was studied to reduce the semantic gap. PRF, as well as interactive RF, were presented with the existing issues.

After the literature review, special attention was paid to finding the most suitable feature selection and representation method to develop a CBIR system to balance the trade-off and be robust to image degradations. Therefore, we collected and study the evaluation measures which will be described in Chapter 3, and then gave our attention to study image features by developing retrieval a system which is included in Chapter 4.

Chapter 3

Datasets and Evaluation Settings

Chapter Organisation

This chapter of the thesis describes the datasets, evaluation measures and methods and parameter settings used in this research work. Section 3.1 represents all the datasets that have been used throughout this research. Section 3.2 and section 3.3 explains the evaluation measures and methodology respectively that were used to evaluate the performance of the content-based image retrieval (CBIR) system. Parameter settings of CBIR system and relevance feedback (RF) system are presented in Section 3.4. The chapter summary and conclusions are included in Section 3.6.

All the details of evaluation matrices, methods, datasets and settings are put together in this chapter because all these things shared across many experiments in many evaluations in all the chapters. Therefore, it is sensible to describe everything in the beginning and referring them when required without repeating all the explanations again in each chapter.

3.1 Datasets

Different standard freely available datasets are used for evaluation purpose and all the datasets are described in this chapter.

Figure 3.1: Example images from the Wang dataset.

3.1.1 Wang Dataset

The Wang dataset [Li and Wang, 2003] of 1000 images is used for both evaluation of the system and comparison. The Wang dataset is a professionally categorised image database and has been widely used to evaluate the performance of CBIR from the past to present [Takala et al., 2005, Hiremath and Pujari, 2007a, Yuan et al., 2011b,Saad et al., 2011,Mansoori et al., 2013,Li et al., 2000,Chen and Wang, 2002,Hiremath and Pujari, 2008,Banerjee et al., 2009,Chowdhury et al., 2012,Hiwale et al., 2015,Lin et al., 2009,Douik et al., 2016,Yang et al., 2015a]. The Wang dataset is a subset of manually selected images from the Corel image database that has been used previously in CBIR as a standard dataset for evaluation purposes. The dataset provides a baseline for comparison with other independently developed and tested approaches. It consists of 10 classes with 100 images in each, namely African people and villages, beaches, buildings, buses, dinosaurs, elephants, flowers, horses, mountains and glaciers, and food. These images are JPEG files with a resolution of 384x256 or 256x384. The type of images contained in the Wang dataset are shown in figure 3.1.

Figure 3.2: Example images from the Oliva and Torralba dataset.

3.1.2 Oliva and Torralba Dataset

For further validation the Oliva and Torralba dataset [Oliva and Torralba, 2001, Gokalp and Aksoy, 2007] is used. It includes 2688 images classified into eight categories, namely coast and beach (360), open country (410), forest (328), mountain (374), highway (260), street (292), city centre (308) and tall buildings (356). These images are JPEG files with a resolution of 256x256. As these datasets are well classified, it was possible to quantitatively evaluate and compare the performance. The type of images contained in the Oliva and Torralba dataset are shown in figure 3.2.

3.1.3 Caltech 256 Dataset

Caltech-256 [Griffin et al., 2007] is an extension of Caltech-101. It has 256 object classes with 30522 images and the smallest class size is 80. Most images are medium resolution, about 300x300 pixels, and of JPEG type. This has higher intra-class variability and higher object location variability than in Caltech-101. Moreover, the objects are not aligned within each class. This is one of the most diverse object databases available today and this has been used in image retrieval evaluation in past years [Prasad and Leung, 2010, Huang et al., 2016, Yang et al., 2015a, Yoon et al., 2014, Silva et al., 2013]. This dataset has more clutter categories as well. The type of images contained in the Clatech256 dataset are shown in figure 3.3.

Figure 3.3: Example images from the Caltech-256 object dataset.

3.1.4 MIR Flickr 25000 Dataset

MIR Flickr25000 dataset [Huiskes and Lew, 2008] is an image collection consisting of 25K images that were downloaded from the social photography site Flickr.com. This database is diverse containing a wide range of objects and scenes. Subsets of this dataset has been used past few years for CBIR evaluation [Xie et al., 2013,Yuan et al., 2013b, Bucak et al., 2011]. These images are representative of a generic domain and provide user tags associated to the images. This dataset has 38 labels including 24 pre-annotations (potential labels) (sky, clouds, water, sea, river, lake, people, portrait, male, female, baby, night, plant life, tree, flower, animals, dog, bird, structures, sunset, indoor, transport, car, people) and 14 regular annotations (relevant labels) (clouds, sea, lake, river, night, tree, flower, dog, bird, car, baby, female, male, portrait) with 24 classes. Images are of different sizes but no larger than 500x500 pixels. The type of images contained in the Flickr dataset are shown in figure 3.4.

3.1.5 Corel 83 classes Dataset

In order to compare our PRF method with other systems that used PRF we conducted another set of experiments on a subset of images from the Corel Photo Gallery [Tao et al., 2007] for which we have results from earlier systems. This dataset has been used from past to present for CBIR and RF evaluations [Tao

Figure 3.4: Example images from the Flickr dataset.

et al., 2006, Zhang et al., 2012a, Li et al., 2006, Tao et al., 2007, Bian and Tao, 2010, Zhang et al., 2014, Zhang et al., 2012b, Hoi et al., 2008, Zhang et al., 2012c]. This dataset is comparably larger and it consists of about 12K images which are classified into 83 semantic classes (castle, lighthouse, modern, sculpture, drinks, fitness, aviation, balloon, bob, bonsai, car, card, decoys, dish, doll, door, Easter egg, flags, mask, mineral, molecular, orbits, ship, steam engine, train, cat, dog, foliage, mushroom, autumn, cloud, firework, forests, ice-burgs, indoor, night, rock form, rural, sunset, waterfall, waves, ski, African, beach, buildings, bus, dinosaurs, elephant, flower, horse, mountain, food, butterfly, cat, cougar, deer, eagle, fish, fox, goat, leopard, lion, lizard, nests, owls, whales (PORP), primates, rhino, tiger, wolf, women, some different arts categories and texture categories) with about 100 images in each. These images are JPEG files with a resolution of 120x80 or 80x120. Even though many images containing the same semantic content are distributed across different Corel categories, we used the same set as it is easy for comparison with the PRF systems in the literature. The type of images contained in the Corel dataset are shown in figure 3.5.

Figure 3.5: Example images from the Corel dataset.

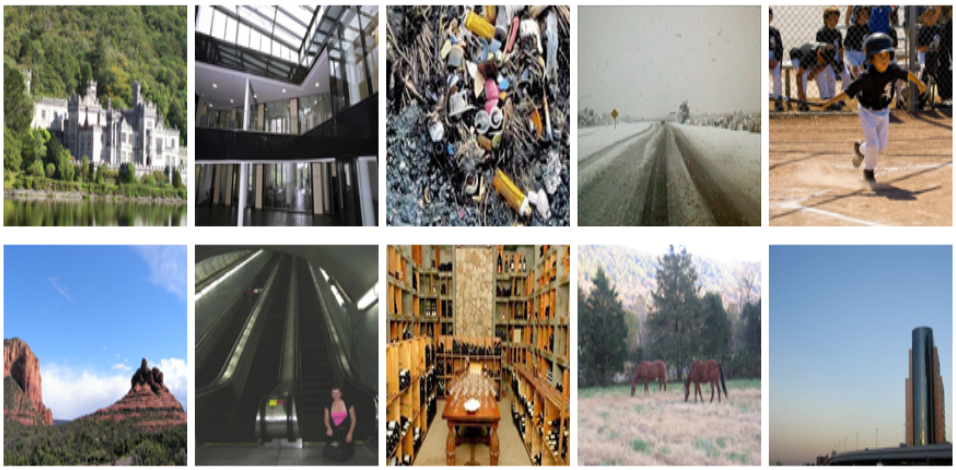


Figure 3.6: Example images from the SUN dataset.

3.1.6 SUN Dataset

SUN dataset [Xiao et al., 2010] is a comprehensive collection of annotated images covering a large variety of environmental scenes, places and objects. It has around 108K images manually annotated into 397 scene categories. Each category contains at least 100 images. These images are mainly categorised into three categories, namely indoor, outdoor natural and outdoor man-made. These images are in JPEG format and are different sizes but are not smaller than 200x200. The type of images contained in the SUN dataset are shown in figure 3.6.

3.2 Evaluation Measures

Note that when evaluating CBIR performances using all these measures, images were considered as correct matches if the retrieved images are belong to the same semantic class of the query image in normal CBIR, PRF and simulated RF. For explicit RF user experiments, relevance information is obtained from the users.

3.2.1 Precision and Recall

Precision and recall are the most common evaluation measures in information retrieval and those measures are used to evaluate the proposed CBIR system.

Recall is defined as the fraction of documents that are relevant to the given query that are successfully retrieved and it is defined as in equation 3.1:

$$Recall = \frac{|Relevant\ images \cap Retrieved\ images|}{|Relevant\ images|} \tag{3.1}$$

hence, measuring the ability of a system to present all relevant items.

Precision is defined as the fraction of retrieved images in a result list that are relevant to a given query and it is defined as in equation 3.2:

$$Precision = \frac{|Relevant\ images \cap Retrieved\ images|}{|Retrieved\ images|} \tag{3.2}$$

Although both measures give a good indication of system performance, they are insufficient if they are just considered alone. A system can achieve higher recall by providing larger output to the user. The system which achieves higher recall, may have really low precision On the other hand, higher precision can be achieve by providing fewer top-ranked images if the system has high early precision. This system will achieve higher precision but with lower recall. Some users prefer early precision while others search for more relevant ones. Therefore, systems always try to balance between these two.

A Precision-Recall curve is used to demonstrate system behaviour with respect to both precision and recall.

3.2.2 Precision @ n

Specifically, the average precision (AP) at n is used to compare with other systems. Here n specifies the length of the result list.

Average Precision for each class $P(c)$ is defined as equation 3.3. Here, c is the class index ($1 \leq c \leq C$) and C is the number of classes in the dataset.

$$P(c) = \frac{1}{N} \sum_{i=1}^{N} p(i) \tag{3.3}$$

Where $p(i)$ is the average precision of i^{th} query image which can be calculated using equation 3.2 and N (or n)is the number of images used for evaluation.

Finally Mean Average Precision (MAP) is calculated using equation 3.4 to calculate the overall precision.

$$MAP = \frac{1}{C} \sum_{j=1}^{c} P(j) \tag{3.4}$$

Where $P(j)$ is the average precision of j^{th} image class and C is the number of classed in the dataset.

AP@20, AP@50 and AP@100 are calculated for each class and MAP of all the classes are considered on Wang, Oliva and Torralba and Corel datasets to compare with existing systems. Images are considered to be correct matches if they are in the same class as the query image for performance evaluation of the proposed CBIR system, PRF and simulated user RF approaches. In real user feedback, images are considered as correct matches if the user selects them as relevant.

3.2.3 R-Precision

This is a special case for a precision value at a certain cut-off. R-precision is the precision after R documents have been retrieved, where R is the number of relevant images for the topic. This can be calculated using the same equation 3.2 and the number of images used for evaluation is R here.

Higher values can easily be achieved for AP@20 and AP@100 if the class sizes are bigger and the classes have many overlaps. In this case, there is higher possi-

bility of getting correct images at the top of the list itself. Therefore, R-precision is used to evaluate those kind of datasets.

3.2.4 Rank of the first relevant image

The simplest measure based on rank is the rank of the first relevant image. This considers the position of the first relevant image to the query from the retrieved list. As this is a poor measure, this was not used for evaluate the retrieval quality.

However, a modified version of this was very useful in evaluating the system's robustness to different image alterations. In the robustness experiment we considered the rank of an unmodified image when the modified image was given as a query image.

3.2.5 Confusion matrix

Each column of the matrix represents the instances in a predicted class while each row represents the instances in an actual class. All correct guesses are located in the diagonal of the table, so it's easy to visually inspect the table for errors, as they will be represented by values outside the diagonal. Therefore, confusion matrix is also used to visually summarise the performance further.

3.2.6 Average Normalised Modified Retrieval Rank (ANMRR)

An Average Normalised Modified Retrieval Rank (ANMRR) [Manjunath et al., 2001] measure is used to counter the bias introduced by different sizes of ground truth sets. The ANMRR is used in the MPEG-7 standardisation process to quantitatively compare the retrieval accuracy. Ranking information is an important issue for retrieval performance. This ANMRR measure incorporates precision, recall measures and rank information. This value is defined as follows:

Consider a query q

$NG(q)$: The number of ground truth images for a query q.

$K(q)$: The top-ranked retrieval results for query q,

where $K(q) = min(4 * NG(q), 2 * maxNG(q))$ as size of ground truth set is normally unequal.

$Rank(k)$: rank of a ground truth image k in retrieval results.

$Rank(k)$is defined as in 3.5:

$$Rank(k) = \begin{cases} rank(k), & \text{if } rank(k) \leq K(q) \\ 1.25 * K(q), & \text{otherwise} \end{cases} \tag{3.5}$$

Average rank $AVR(q)$ of the images for query q is defined as in equation 3.6:

$$AVR(q) = \frac{1}{NG(q)} \sum_{k=1}^{NG(q)} Rank(k) \tag{3.6}$$

To minimise the influence in $NG(q)$,a modified retrieval rank (MRR) is defined as in equation 3.7:

$$MRR(q) = AVR(q) - 0.5 * \Big[1 - NG(q)\Big] \tag{3.7}$$

Still, the upper bound depends on $NG(q)$. To normalise the value, nNrmalised Modified Retrieval Rank (NMRR) is defined as in equation 3.8:

$$NMRR(q) = \frac{MRR(q)}{1.25 * K(q) - 0.5 * \Big[1 + NG(q)\Big]} \tag{3.8}$$

$NMRR(q)$ has values between 0 (indicating whole ground truth found) and 1 (indicating nothing found) irrespective of the size of ground truth.

Finally, consider the average NMRR of all queries using equation 3.9:

$$NMRR(q) = \frac{1}{NQ} \sum_{q=1}^{NQ)} NMRR(q) \tag{3.9}$$

3.3 Evaluation Methods

The first pass retrieval result list is used to evaluate the proposed preliminary system and the initial CBIR-ISIG system.

Three evaluation methods are used to empirically evaluate the RF system:

Full Ranking - the entire set of retrieved signatures is re-ranked using RF.

Freezing - the initial top-ranked signatures are frozen and the RF system is

used to re-rank the remaining signatures.

Residual Ranking - the top-ranked signatures are removed from the ranked set after being used to train the RF system.

Residual ranks and frozen ranks are useful methods to explore the effectiveness of RF methods when using real user feedback. We use these methods for the evaluation of PRF, simulated user RF and interactive RF as well, in order to factor out the effect of RF on the results in the evaluation.

3.4 Parameter Setting

Different parameter settings were used to achieve better retrieval performance in the proposed preliminary image retrieval systems, signature based image retrieval system and the proposed RF approach. These parameters are described in detail in following chapters.

3.4.1 Normal CBIR-ISIG System

Image Subdivision for Feature Extraction:
This research uses image decomposition to generate bag of words representation. Different image decomposition methods are compared in this research so we can select the best decomposition method. All the decomposition methods are compared under same experimental setting (dataset, signature size).

Vocabulary Size of the Bag of Words Approach:
The main representation of this thesis is based on the approach of bags of words. In information retrieval the texts are represented by sets of words from a vocabulary built from the corpus and size of text vocabulary is predefined by the corpus. However, in image retrieval the size of visual vocabulary is obtained by clustering methods. In here vocabularies must be made for each descriptor from the extracted features. The vocabulary size may have an impact on the effectiveness. A small vocabulary may provide lower accuracy due to the lack discriminative power since some feature vectors may be assigned into the same cluster even if they are not similar to each other. On the other hand, a large vocabulary may not leads to higher accuracy as it is sensitive to quantization errors and acquires extra processing overhead as well. Therefore, experimental evaluations were done over a variety of vocabulary sizes and that vocabulary size depends on the dataset

size. Thus, it is necessary to select an appropriate vocabulary size that produces the best performance.

Term Statistics of the Vocabulary:
To achieve better retrieval quality, term weighting is applied on Bag of Features (BoF). Weighting helps to improve precision and recall. This is the motivation behind Term Frequency-Inverse Document Frequency (TF-IDF) technique and log-likelihood used in text retrieval. Term weights can be computed in many different ways, like Term Frequency (TF), TF-IDF, log-likelihood and Okapi BM25. Experiments were carried out with mostly used term statistics on image datasets. Finally, term statistics were not used for the current datasets by considering the trade-off between retrieval quality and search time, but term statistics are essential to address very large datasets.

The term statics which were experimented are TF-IDF [Salton et al., 1975], Okapi BM25 [Manning et al., 2008] and log-likelihood [Chappell, 2015]. Their definitions are given bellow.

- **TF-IDF**

$$TF - IDF(t) = tf(t) * log\frac{N}{1 + df(t)} \tag{3.10}$$

  Where,

  - $tf(t)$ (Term frequency): the number of times this term appears in the document.
  - N (Document count): the number of documents in the collection.
  - $df(t)$ (Document frequency): the number of documents this term appears in.

- **Okapi BM25**

$$BM25(t) = \left(\frac{tf(t) * (K_1 + 1)}{tf(t) + K_1(1 - b - b\frac{D}{avg})} + \delta \right) * log\frac{N - df(t) + 0.5}{df(t) + 0.5} \tag{3.11}$$

  Where,

- $tf(t)$ (Term frequency): the number of times this term appears in the document.
- N (Document count): the number of documents in the collection.
- $df(t)$ (Document frequency): the number of documents this term appears in.
- D (Document length): Length (in terms) of the current document.
- avg (Average document length): Average length (in terms) of documents in this collection.
- K_1 and b (Tuning parameter): Constant values

- **Log-likelihood**

$$LL(t) = (tf(t) - 0.5) * log(\frac{tf(t}{D} * \frac{C}{tcf(t)}) \tag{3.12}$$

- $tf(t)$ (Term frequency): the number of times this term appears in the document.
- $tcf(t)$ (Term collection frequency): the number of times this term appears in the collection.
- D (Document length): Length (in terms) of the current document.
- C (Collection length): Length (in terms) of every document in the collection.

Binary Image Signature Size for Image Representation:
Signature size has a trade-off between accuracy and efficiency. Larger signatures provide a higher quality representation of the underlying document than smaller signatures but require more processing time to create, more processing time to search and more memory and disk space to store. Different scenarios may favour different trade-offs, so it is worthwhile to consider how different signature sizes affect search quality and processing time. Thus, next we evaluated how signature size affects precision.

3.4.2 CBIR-ISIG System with Relevance Feedback

Re-rank List Size in Relevance Feedback:
This proposed RF approach is different from typical RF approaches. The full database is searched only the first time. Top-ranked signatures are then already resident in memory after the initial search. This RF approach uses subsets of the rank list to re-rank after RF, namely the "Re-rank list", to increase the computational efficiency in subsequent search processes. This list size is significantly lower than the size of the image dataset. Experiments were carried out to see the effect of re-rank list size on the performance of RF approach.

Feedback Sample Size to Generate Feedback Signature:
When RF approach is applied to CBIR system the feedback, that input to the system as positive and negative feedback affects the performance of RF to the next iteration. This feedback size is considered as feedback sample size (feedback sample size << Re-rank List Size). Experiments were carried out to see the effect of feedback sample size on the performance of RF approach.

Scaling Factor to Generate Feedback Signature :
This research is proposes a rank-based RF approach to improve retrieval quality of the CBIR. Therefore, scaling factor was essential to include the rank information when generating feedback signature from the PRF, simulated user RF and real user feedback. Different scaling factors were tested using PRF approach. This scaling factor ensures that the signatures that are closer to the top of the retrieved list contribute more to the new feedback signature (which is used as query signature in next run). After several experiments a scaling factor was selected which was used in this thesis.

Value w in Scaling Factor:
w is the decay factor in the scaling factor. In particular, w controls the granularity of the image similarity. Larger the value of w, the more RF promotes images that are visually similar to those selected for RF. Several experiments were carried out to determine a value for w.

3.5 Results

The above mentioned datasets are used to evaluated the proposed CBIR performance on different occasions. These evaluation measures and methodologies were used appropriately to empirically evaluate system performance. How the experiments were carried out and the analysis of results are included in each chapter.

3.6 Chapter Summary and Conclusions

This chapter gives an overview of the datasets, evaluation measures and methodologies and parameter setting used in the evaluations in each chapter of this thesis.

Chapter 4

Preliminary Work - Simple but Effective Techniques to Improve Content-Based Image Retrieval

Chapter Organisation

This chapter presents the preliminary work done in this research and the findings during the early research work. Section 4.2 briefs the selected features for the system. Section 4.3 explains the proposed system based on a multi-level searching approach, related work on that subject, performance evaluation, results and the conclusion drawn from the analysis of results. Section 4.4 explains the proposed system using late-feature fusion, related work on that subject, performance evaluation, results and the conclusion drawn from the analysis of results. The chapter summary and conclusion are included in Section 4.5. The original contributions discussed in this chapter resulted in publication (ii) and (iv) in List of Publications. When developing the main system in Chapter 5, only appropriate features were selected from the features studied in this chapter.

4.1 Introduction

The development of the Internet and the increased availability of image capturing devices have enabled collections of digital images to grow at a fast pace in recent years and to become more diverse. This has created an ever-growing need for efficient and effective image browsing, searching and retrieval tools. Content-Based Image Retrieval (CBIR) for a general image database is a highly challenging problem. A number of representations and searching techniques have been developed for general-purpose search engines.

Some key issues that need to be answered when developing CBIR systems are as follows: choosing the features that can be used to extract image properties, the way of extracting and presenting image features, and finally, determining the similarity measure to retrieve visually similar images. These issues have been addressed in several ways and a number of techniques have been proposed in the literature [Patvardhan et al., 2013, Hiremath and Pujari, 2007a, Saad et al., 2011, Mansoori et al., 2013, Takala et al., 2005, Yuan et al., 2011b, Li et al., 2000, Ruikar and Kabade, 2016, Shrivastava et al., 2015, Mariam and R, 2015].

Image features are essential for CBIR, as image content needs to be extracted for retrieval. Colour, texture and shape features have been used to retrieve visually similar images from an image database. Most of the systems have used one or two features , while few systems have used all the features [Hiremath and Pujari, 2007a, Ruikar and Kabade, 2016, Shrivastava et al., 2015, Mariam and R, 2015]. Features that are effective in terms of differentiating images have to be chosen according to the type of the dataset. All the features may be useful for general image collection as it is heterogeneous (i.e. there may be colour images, natural real-world images, as well as images of objects). One feature or two will not be enough to describe and distinguish between images.

When we consider the image features, those features have different abilities to retrieve images. Image features must be selected according to several factors such as their ability to search relevant images, computational complexity and the application that is going to be used. Moreover, feature inter-dependency must be considered if these features are used together (for feature fusion).

This chapter describes image features selected for the content-based image

retrieval in this thesis and some other preliminary work done in this research before going into the main research objective and the findings during the early research work. The related work to the each task is described separately in Section 4.3.1 and 4.4.1. Research findings are summarised in Section 4.3.6 and 4.4.8.

4.2 Image Features

Feature extraction plays a major role in CBIR. The comparison of various defined techniques indicates that using a single feature for image retrieval is not an adequate solution. Multi-feature representation for image retrieval was therefore necessary and an approach was proposed, which uses a combination of low-level features such as colour, texture, shape and GIST descriptors. The main reason to select these descriptors was to address large heterogeneous databases of images. The following will provide a brief on the features with their feature descriptors, which are critical for accurate retrieval.

- **COLOUR - COLOUR HISTOGRAM (CH)**

  The advantages of Colour Histograms are the efficiency and insensitivity to small changes in camera view-point. This research used colour patterns, which was generated by G.Qiu [Qiu, 2002] in 2002, as it achieved better performance. There are many colour space models but YCbCr colour space is much more suitable for human visual systems and video systems. G.Qiu generated a codebook of 64, 128 and 256 for chromatic and achromatic colour patterns. This colour pattern was defined as 'the spatial and spectral characteristic of a small block of pixels in a colour image' [Qiu, 2002]. Here they considered small block size as 4x4. If an image is mxn, they generate (mxn) / 16 blocks.

  As demonstrated in Figure 4.1, the visual appearance of a small image block is modelled by three components [Qiu, 2002]. S - Stimulus Strength, P - Spatial Pattern, C - Colour Pattern. The visual appearance of a small image block was modelled by three components: the stimulus strength (S), the achromatic spatial pattern (P) and the chromatic spatial pattern (C) [Qiu, 2002].

P channel captures the achromatic spatial pattern, while C channel captures the chromatic spatial pattern of the input coloured image. G.Qiu [Qiu, 2002] generated codebooks for achromatic and chromatic patterns separately by using a frequency sensitive competitive learning algorithm because it is insensitive to initial selection of codewords and can be used more efficiently than those designed by other vector quantisation methods. To generate a codebook, more than 15 million training samples were used and each of them was trained for 20 times.

Histogram generation is as follows and a 64 dimensional colour histogram was generated. This research used 64 codewords; 32 Achromatic Spatial Patterns (ASP) and 32 Chromatic Spatial Patterns (CSP). Though authors were using them separately, they were combined as one single vector in our research.

- First convert image from RGB space to YCbCr.
- Then Image was subdivided into non-overlapping 4x4 size blocks.
- Strength (S) of the block (mean) was calculated. Let

$$Y = y(i, j), \qquad i, j = 0, 1, 2, 3$$

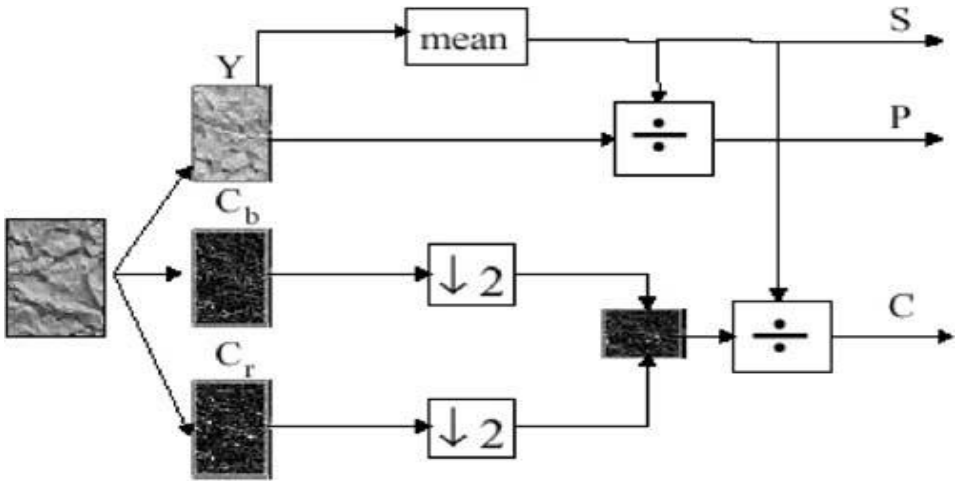


Figure 4.1: Coloured Pattern Appearance Model (CPAM).

be the 4x4 Y image block. Then stimulus strength (S) was calculated by equation 4.1.

$$S = \frac{1}{16}\sum_{i=1}^{3}\sum_{j=1}^{3} y(i,j) \tag{4.1}$$

– Each pixel in the block was divided by the mean S to generate pattern vector ASP.

$$ASP = \{asp(i,j), \qquad i,j = 0,1,2,3\} \tag{4.2}$$

where

$$asp(i,j) = \frac{y(i,j)}{S}$$

– CSPvector was formed by sub-sampling the two chromatic channels Cb and Cr to gain a single vector

$$Cb = \{c_b(i,j), \qquad i,j = 0,1,2,3\}$$

$$Cr = \{c_r(i,j), \qquad i,j = 0,1,2,3\}$$

Sub-sampled signal

$$Cb; \quad SCb = \{sc_b(k,l), \qquad k,l = 0,1\} \tag{4.3}$$

and

$$Cr; \quad SCr = \{sc_r(k,l), \qquad k,l = 0,1\} \tag{4.4}$$

were obtained from equation 4.5 and equation 4.6

$$sc_b = \frac{1}{4S}\sum_{i=0}^{1}\sum_{j=0}^{1} c_b(2k+i, 2l+j) \tag{4.5}$$

$$sc_r = \frac{1}{4S}\sum_{i=0}^{1}\sum_{j=0}^{1} c_r(2k+i, 2l+j) \tag{4.6}$$

- Then the CSP vector was generated

$$CSP = \{csp(k), \qquad k = 0, 1, , 7\}$$

  by concatenating SCb and SCr
- Pattern vector ASP and vector CSP were provided and a related index was found using 32 codewords for both pattern and colour. Two 32 dimension vectors were generated to store values according to indices as $histP$ and $histC$ for pattern and colour respectively. The value one was added to those indices in new $histP$ and $histC$ accordingly.
- The above steps were repeated until all the image blocks were finished. In each step the nearest index was found and thw count was increased by one for that specific index value for both $histP$ and $histC$.
- Finally, $histP$ and $histC$ were normalised by dividing the sum of all the values in the vector and then they were combined to generate one vector.
- A 64 dimensional colour histogram was generated.

- **COLOUR - COLOUR COHERENCE VECTOR (CCV)**

  The advantage of a Colour Coherence Vector is that it includes spatial information, unlike a colour histogram. This research referred to the CCV proposed by Greg Pass et.al [Pass et al., 1996]. Here, each pixel in a colour bucket was classified as coherent or incoherent, based on whether it is a part of a large similarity coloured region or not [Pass et al., 1996]. The first image was blurred and then the threshold was applied as 127 to all the components and pixel values higher than 127 were replaced by 1 and others by 0.

  The eight colour components used were (R=1, G=2, B=3 RG=4 RB=5 GB =6, White=7 and Black=8). Then each pixel was classified into these components and further classified the region by colour. After that, each region was checked and if the region was bigger than 1% of the image size, that region was considered as an area filled with coherent pixels and an area filled with incoherent pixels otherwise. The coherent and incoherent pixels

for each colour component were counted and used as the feature vector. The feature vector was 16 dimensional.

- **COLOUR - COLOUR MOMENTS (CM)**

  Colour Moments are simple and describes the colour distribution. Therefore, colour moments were adapted in this research to describe colour. The first order original moment, second order central moment and third order central moment were calculated for each colour component in an RGB image. Finally, a nine dimensional feature vector was generated to describe colour distribution.

- **TEXTURE - GABOR FEATURES (GABOR)**

  Gabor Features are well known and widely used method for texture retrieval. This research also used it to extract image features. Texture features were extracted using following equations. Four scales and six orientations were used to extract the feature as it was the best selection for our experiments. Rotation and scale invariance property was achieved by the simple circular shift operation proposed in [Rahmana et al., 2011]. Finally, the feature vector was generated by the calculating mean and standard deviation of each filer output. Here the feature vector was 48 dimensional.

- **TEXTURE - WAVELET TRANSFORM (DWT)**

  Discrete Wavelet Transforms provide a good multi-resolution tool for texture description and it is allowed to represent texture at the most suitable texture having various spatial resolution. This research used the simplest form of wavelet decomposition. Images were converted into YUV colour space and Daubechies wavelets (db20) were used because it was better than Haar wavelets for general purpose images. In this research, decomposition was done only up to three levels and each time the low frequency sub band (LL) was decomposed. Then, mean and standard deviation of each HL (vertical edge features) and LH (horizontal edge features) bands were computed. Rotation invariance property was achieved by the simple operation proposed in [Manthalkar et al., 2003]. Mean and standard deviation of each HL and LH band were summed and divided by two, in correspondence to its decom-

position level to achieve rotation invariance and final vector was generated with 42 dimensions. Here, HH sub band was neglected because it contains the majority of noise in the image.

- **TEXTURE - EDGE ORIENTATION HISTOGRAM (EHD)**

  The Edge Orientation Histogram (Edge Histogram Descriptor) captures the spatial distribution of edges and this helps to extract different textures. It is translation invariant and robust to partial occlusion and local disturbances in the image. Moreover, edge orientation histogram [Agarwal et al., 2013] computation is easy and effective and has been used in many applications to extract texture feature [Chaudhary and Upadhyay, 2014, Saavedra, 2014]. We selected this feature for our CBIR system. To detect edges, the image was convolved with five sobel operators (horizontal edge, vertical edge, 45-degree edge, 135-degree edge, and non-directional edge) which contain 3x3 kernels and found the maximum of Sobel gradient. The binary image was generated using the canny edge detector. The binary image was multiplied with the types of orientation detected above. The image was partitioned into 16 non-overlapping blocks. Finally, an 80 dimensional feature vector was generated by combining all the histograms.

- **SHAPE - GENERIC FOURIER DESCRIPTORS (GFD)**

  Generic Fourier Descriptor [Zhang and Lu, 2002] is region-based method and it is suitable for general image retrieval and its translation, rotation, scale invariant and robust to noise and occlusion. In this method, image in polar space was considered as a two dimensional rectangular image in Cartesian space. Four radial frequencies and 15 angular frequencies were used to generate 60 dimensional GFD.

- **SHAPE - MOMENT INVARIANT (MI)**

  Moment Invariant [Rai et al., 2011] is an invariant feature and is widely used for shape retrieval tasks. It provides a compact representation of the pixel distribution of a shape image. Only seven invariant moments were used because higher order moments are sensitive to noise.

- **Other - GIST**

Though GIST is a global descriptor, it has achieved very good performance in literature. Torralba [Torralba et al., 2008a] has been used this for 80 million tiny images dataset. This research adapted the code developed for the SUN database to represent images [Xiao et al., 2010]. It has differences from the original method. It uses RGB colour space with 20 filters in three scales (as [8 8 4]) ended up as a 960 ((4x4) x3 x20) dimensional feature vector. We used gray scale with 8 orientations and 3 scales ((4x4) 8 24 = 384).

4.3 An Effective Content Based Image Retrieval System Based on Multi-Level Searching

The objective of this research work was to develop a simple CBIR system in order to achieve adequate precision in image classification and retrieval using the global representation. According to the literature, it is difficult to achieve good performance using the global representation. However, we contradicted that statement and showed that even the global representation can be used to achieve better retrieval performance through the addition of new techniques. As a single feature was not adequate for CBIR of a general dataset, colour texture and shape features were used. Each feature was represented by a single feature vector for each image. A multi-level sequential searching technique was used for image searching. Empirical evaluation was performed on two subsets of the standard Corel dataset and validated the performance of this method against other independently evaluated methods [Chathurani et al., 2015b]. This proposed approach will be described in a more detailed manner in this section.

4.3.1 Background Work

There are mainly two ways of image representation in CBIR, namely local representation and global representation. Only a few systems have used global representation [Saad et al., 2011], while most of the systems have used local representation [Hiremath and Pujari, 2007a, Mansoori et al., 2013, Takala et al., 2005] because the local representation gives better results than the global representation. But this is not always the case. The local representation is applied to a wide range of CBIR systems and applications to achieve robustness. Local feature extraction relies on the detection of landmark points or the segmentation of the image into regions. However, a precise image segmentation method that can be applied to general image collections has not yet been found. The global representation is simple, as it extracts the features from the full image without subdividing or searching points of interest of the image. This research work shows how better results could be achieved even with global representation. The size of the feature database is

kept reasonably small.

After the representation, a similarity mechanism must be defined. Different similarity measures such as region matching and histogram matching have been used in the local representation [Hiremath and Pujari, 2007a, Yuan et al., 2011b, Li et al., 2000]. However, for the global representation [Saad et al., 2011] regardless of the selected similarity measure, a single-level simple sequential search has been the single choice in the typical CBIR. Querying by colour, texture, shape or one of the combinations of these features has been proposed in several systems [Saad et al., 2011, Mansoori et al., 2013, Yuan et al., 2011b] using single-level sequential searching. In the single-level sequential search, features are fused to generate one feature vector or different feature vectors. Then the features are used in the same level with or without weightings for searching. It seems that multi-level sequential search has not been considered so far, even though it is simple and shows improved retrieval results. To the best of our knowledge multi-level sequential search is yet to be studied to improve the retrieval results.

4.3.2 Image Features

Different low-level features were used as a combination of colour, texture and shape. Well-known Colour Coherence Vector, Colour Histogram and Colour Moments which extract different variations of colour were selected as the colour feature to describe an image. YCbCr and CIEluv colour spaces were used, as those provide a closer match to human perception. Well-known Gabor Wavelet, Discrete Wavelet Transform and Edge Histogram Descriptors were used as texture descriptors in this method. Invariant Moments and Generic Fourier Descriptors were used for shape retrieval.

Feature descriptors were selected for each feature using cross validation. It must be noted that we only considered combinations of feature descriptors that related to a particular feature at a time, which means colour, texture and shape separately as we used colour, texture and shape in three different stages.

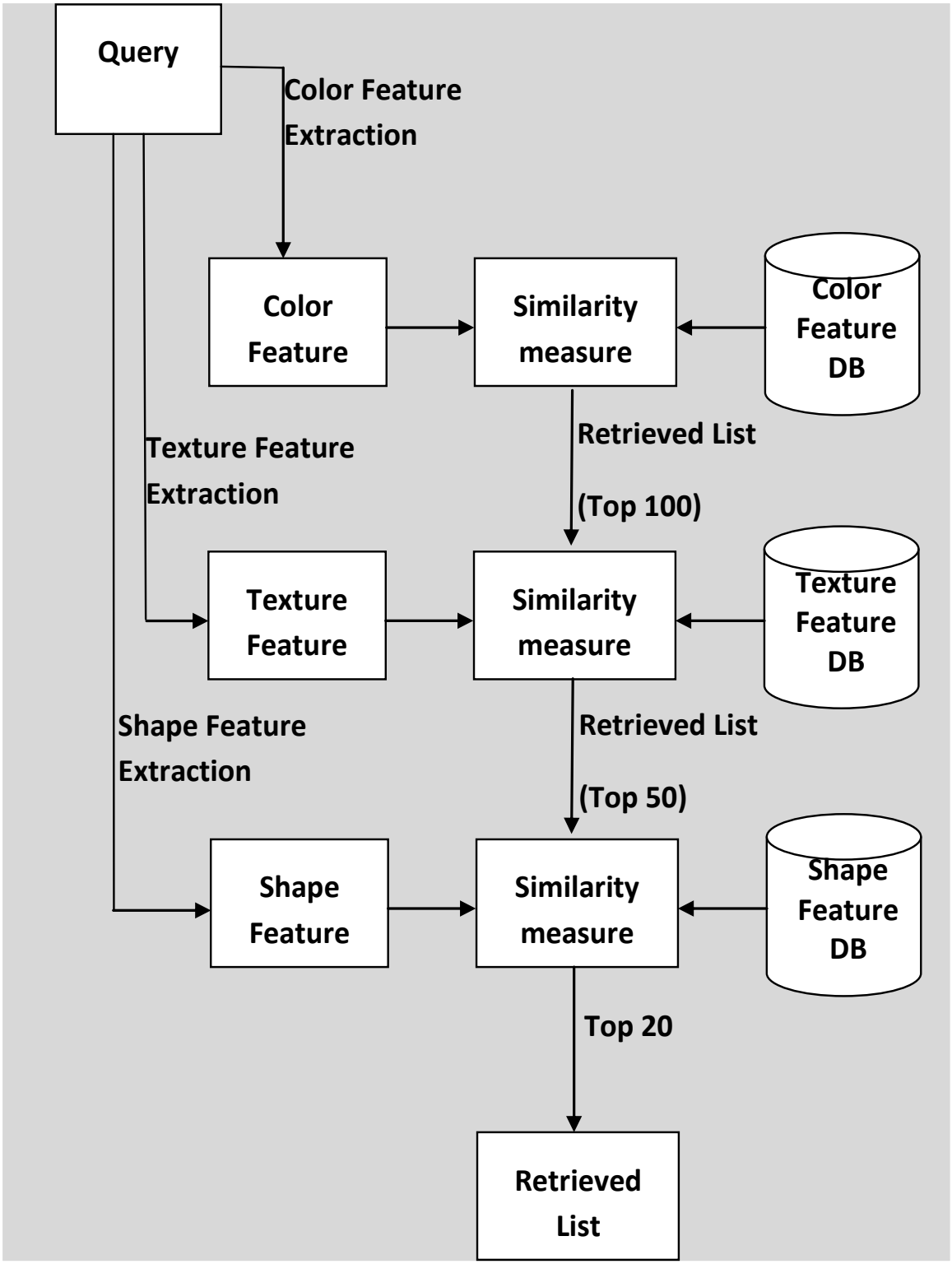


Figure 4.2: An overview of the proposed multi-level sequential searching process.

4.3.3 Feature Representation

The global image representation was selected and the results highlighted that even a global representation can be used to achieve better retrieval performance in CBIR. Features were extracted from the full image for the whole dataset. Three individual feature vectors were used to represent each image; for colour, texture, and shape separately. The feature vectors were normalised to reduce the effect of bias in large values. Finally, three databases were generated for the image database as shown in figure 4.2.

4.3.4 Image Searching

Normally, single-level sequential search mechanisms were used to retrieve the most relevant images from the image database in CBIR systems. Even though multiple feature representation for image retrieval is necessary, performance depends not only on the image features but also on feature representation. There is a possibility that the expected retrieval performance will not be achieved when combining features together because one feature may dominate the entire system performance. It was found that sometimes performance got worse when using simple feature combinations than using a single feature, as an example, when features were overlapped. Feature vectors need to have weak correlation between each other to provide improved performance. Sometimes one feature may be more important than the other features for retrieval.

A novel multi-level sequential searching mechanism (MLSS) was proposed in order to achieve adequate precision in image classification and retrieval using the simple global representation as a solution for the above mentioned problem. With multi-level we refer to the fact that image search was carried out in three levels by selecting appropriate feature order according to the type of the dataset (dataset of objects, colour images, texture images or heterogeneous collection). For example, shape feature is used in the initial stage to search similar images and images were re-ranked using that feature. Now we have retrieved an images list from shape feature and that list is re-ranked using texture feature by considering only a subset of the image list from the prior retrieved list. In the second level we use the same image list which is provided by the shape feature to texture feature, but

re-ranked. Now, a subset of this list is used to re-rank using the colour feature and the final retrieval list is taken after the colour and shown to the user. This example case may be more useful for object detection as we give shape feature the most importance among the three. The purpose of this was to use image features according to their importance to the database. As we used general images, colour, texture, shape order was the most suitable order to search images and it was found after the experimental evaluation.

When a query image was given to the system, the top $N_1(100)$ images were found using the colour feature database in the first stage and this list was fed to the second stage. In the second stage the top N_2 (50) images were selected from the given list (N_1) using the texture feature database. Finally, the list N_2 was fed to the third stage and top N_3 (20) images were selected using the shape feature database and displayed to the user. This system showed the first 20 images to the user. This was the process of searching in the proposed CBIR system and the system was evaluated using two standard datasets. For evaluation purposes, the above values of N_1, N_2, N_3 ($N_1 = 100$, $N_2 = 50$, $N_3 = 20$) and ($N_1 = 200$, $N_2 = 150$, $N_3 = 100$) were used.

This research work was concerned with the effect of multi-level searches on CBIR retrieval quality. An overview of the proposed multi-level searching process is depicted in Figure 4.2 for more information.

4.3.5 Experimental Results

To evaluate the effectiveness of the proposed feature fusion approach, experiments were performed on two general purpose image datasets; Wang and Oliva and Torralba. The detail of these datasets can be found in Section 3.1.1 and Section 3.1.2 in Chapter 3.

Experimental setup

As real users were not interacting with the system, the system was evaluated on classification as most of the other systems [Li et al., 2000, Chen and Wang, 2002, Takala et al., 2005, Hiremath and Pujari, 2007a, Hiremath and Pujari, 2008, Yuan et al., 2011b, Saad et al., 2011, Mansoori et al., 2013] which were used in this work had used classification information for system evaluation. Images were classified

Table 4.1: Average Precision (AP) of the Wang dataset for single-level search and multi-level search (AP @ 20)

Class	Single-Level Search					Multi-Level Search	
	Colour (Col)	Texture (Tex)	Shape	Whole Set No Weight	Whole Set Weighted	Col Tex Shape	Tex Col Shape
Africans	0.65	0.68	0.51	0.60	0.60	0.72	0.71
Beach	0.30	0.39	0.37	0.34	0.36	0.39	0.40
Building	0.39	0.42	0.38	0.36	0.36	0.44	0.43
Bus	0.66	0.73	0.8	0.75	0.79	0.85	0.87
Dinosaur	0.99	0.99	0.99	1.00	1.00	1.00	1.00
Elephant	0.49	0.43	0.52	0.53	0.54	0.56	0.55
Flower	0.76	0.65	0.65	0.74	0.76	0.75	0.73
Horse	0.93	0.89	0.69	0.85	0.85	0.89	0.84
Mountain	0.42	0.39	0.31	0.43	0.44	0.48	0.48
Food	0.57	0.63	0.36	0.52	0.53	0.58	0.58
Average Precision	0.616	0.62	0.558	0.612	0.623	0.666	0.659

into several categories (with similar content) and then retrieval accuracy could be evaluated using this classification data. A retrieved image was considered a correct match if and only if it was in the same category as the query image.

The most common evaluation measure in information retrieval is precision and it is calculated by using equation 3.2. The multi-level sequential searching mechanism was based on average precision by evaluating the top 20 (N=20) and 100 (N=100) retrieval results for comparison. When N=20 (top lists are 100 and 50 consequently), when N=100 (top lists are 200 and 150 consequently).

Confusion matrix for each dataset was calculated to visually summarise the performance further.

The Precision-Recall (PR) curve is a very good measure to evaluate a system. So precision @ recall was calculated. Recall was calculated using equation 3.1

Results

Before comparing the results with other published systems, the proposed system is compared with the normal single-level search and proved that better searching results can be achieved by the multi-level search than the single-level search. Ta-

Table 4.2: Average Precision (AP) of the Oliva and Torralba for single-level search and multi-level search(AP @ 20)

Class	Single-Level Search		Multi-Level Search	
	Whole Set No Weight	Whole Set Weighted	Col Tex Shape	Tex Col Shape
Coast (beach)	0.34	0.35	0.40	0.35
Open country	0.32	0.32	0.45	0.40
Forest	0.25	0.25	0.50	0.45
Mountain	0.28	0.32	0.35	0.40
Highway	0.25	0.27	0.55	0.45
Street	0.32	0.32	0.45	0.45
City centre	0.30	0.31	0.40	0.35
Tall buildings	0.23	0.25	0.35	0.30
Average Precision	0.2863	0.2988	0.4313	0.3938

ble 4.1 and table 4.2 show the self-comparison results for both datasets. Table 4.1 shows average precision for each feature separately, as well as normal feature combination (one vector), feature combination with feature weights for each feature in single-level sequential search and multi-level sequential search on the Wang dataset. Table 4.2 shows average precision for single-level sequential and multi-level sequential search on the Oliva and Torralba dataset. Significance testing was performed using 2-tail t-test [Fay and Proschan, 2010] with p-value$<$0.01 (99% significance level) and p-value$<$0.05 (95% significance level) as it is common in information retrieval. According to the results in table 4.1 performance of multi-level colour, texture, shape retrieval search is significantly better at 99% significance level except texture, colour and shape order (95% significance level). According to the results in table 4.2 colour, texture and shape multi-level search shows significantly better performance compared to the other three methods at 99% significance level.

Confusion matrices for multi-level sequential search for both the datasets are presented in table 4.3 and table 4.4.

Table 4.3: Confusion matrix for Wang dataset

True Class \ Assigned Class	Africans	Beach	Building	Bus	Dinosaur	Elephant	Flower	Horse	Mountain	Food	Total	% Assigned
Africans	16	0	1	0	0	1	0	0	0	2	20	80.00
Beach	1	8	2	1	0	2	0	0	5	1	20	40.00
Building	2	2	10	2	0	1	0	0	2	1	20	50.00
Bus	0	1	1	18	0	0	0	0	0	0	20	90.00
Dinosaur	0	0	0	0	20	0	0	0	0	0	20	100.00
Elephant	2	1	2	0	0	12	0	1	1	1	20	60.00
Flower	2	0	0	0	0	0	16	0	0	2	20	80.00
Horse	0	0	0	0	0	1	0	19	0	0	20	95.00
Mountain	1	4	2	1	0	2	0	0	10	0	20	50.00
Food	3	0	1	1	0	2	1	0	0	12	20	60.00
Total	27	16	19	23	20	22	17	20	18	19	100	70.00

Table 4.4: Confusion matrix for Oliva and Torralba dataset

True Class \ Assigned Class	Beach	Country	Forest	Mountain	Highway	Street	City	Building	Total	% Assigned
Beach	9	4	1	2	2	0	1	1	20	45.00
Country	3	10	2	2	2	1	0	0	20	50.00
Forest	1	3	10	2	0	2	1	1	20	50.00
Mountain	2	2	1	9	2	2	1	1	20	45.00
Highway	2	2	1	1	11	1	1	1	20	55.00
Street	1	0	1	2	1	9	4	2	20	45.00
City	0	2	0	2	1	5	8	2	20	40.00
Building	0	2	1	2	1	3	3	8	20	40.00
Total	18	25	17	22	20	23	19	16	100	46.00

According to the results of table 4.1 and table 4.2 it can be seen that the performance of the Wang dataset is much better than the Oliva and Torralba dataset. The main reason for this may be that the classes of Wang dataset are clearly separable and it is not so in the Oliva and Torralba dataset. As shown in table 4.4, same coloured boxes depict the visually similar images. (Ex: some city centre and tall building images are visually similar to street images. That's why we gain these values in confusion matrix for these classes. This can be checked with the images in that dataset.) If we consider those as correctly classified, the system can achieve better retrieval results.

Figure 4.3 provides the precision-recall curve for Wang dataset. It shows the ability of the proposed searching algorithm to retrieve images. It has around 50 % precision at 50 % recall.

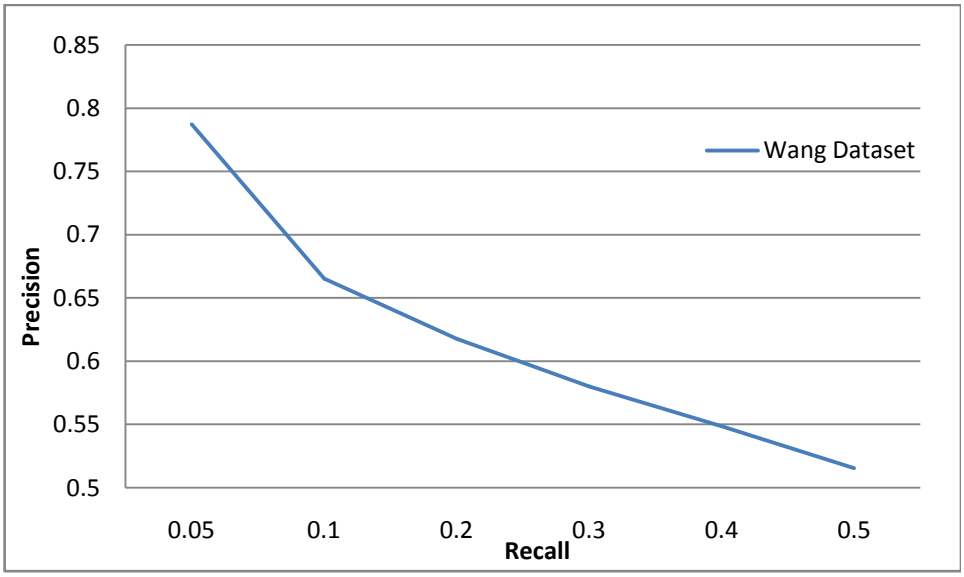


Figure 4.3: Precision-Recall curve for Wang dataset.

The systems that were compared are represented by publication year and the reference in table 4.5 and table 4.6. Table 4.5 shows the results of the proposed CBIR system compared with the other systems (average precision for the top 20) and it shows that the proposed system generates better results when using this approach. As the other systems only provide the mean average precision of each class but not the results for each query, significant test cannot be computed

Table 4.5: Average Precision (AP) of each class along with the whole dataset for the Wang dataset compared with the performance of the systems in the literature (AP @ 20)

Class	2005 [Takala et al., 2005]	2007 [Hiremath and Pujari, 2007a]	2011 [Yuan et al., 2011b]	2011 [Saad et al., 2011]	2013 [Mansoori et al., 2013]	MLSS Our Method
Africans	0.23	0.48	0.57	**0.90**	0.70	0.72
Beach	0.23	0.34	**0.58**	0.38	0.28	0.39
Building	0.23	0.36	0.43	**0.72**	0.56	0.44
Bus	0.23	0.61	**0.93**	0.49	0.84	0.85
Dinosaur	0.23	0.95	0.98	**1.00**	0.81	**1.00**
Elephant	0.23	0.48	**0.58**	0.39	**0.58**	0.56
Flower	0.23	0.61	**0.83**	0.56	0.55	0.75
Horse	0.23	0.74	0.68	0.87	0.87	**0.89**
Mountain	0.23	0.42	0.46	0.45	**0.48**	**0.48**
Food	0.23	0.50	0.53	**0.87**	0.66	0.58
Average Precision	0.23	0.549	0.657	0.663	0.633	**0.666**

but averages are higher. In addition, table 4.6 shows the results of the proposed system compared with other systems (average precision for the top 100, this can be considered as recall as each dataset has 100 images). Here total average precision reaches the highest in the proposed system for the top 20 and average performance for the top 100.

$N_3 = 20$ and 100 are used for ease of comparison with existing systems as they have evaluated AP @ 20 and AP @ 100. The order of features to be used in multi-level sequential searching is selected from the experimental results. The Wang dataset was tested for all the possible combinations (9) and found that the order of colour (C), texture (T), shape (S) and T, C, S consecutively in each step gave the best results among those combinations. They were put in this order because the dataset is general. Then the Oliva and Torralba dataset was tested for these combinations and gained similar output. By considering the results in the last two columns of table 4.1 and table 4.2 the C, T, S feature order is used for searching

Table 4.6: Average Precision (AP) of each class along with the whole dataset for the Wang dataset compared with the performance of the systems in the literature (AP @ 100)

Class	2000 [Li et al., 2000]	2002 [Chen and Wang, 2002]	2008 [Hiremath and Pujari, 2008]	2008 [Hiremath and Pujari, 2008]	MLSS Our Method
Africans	0.48	0.47	0.40	0.48	**0.49**
Beach	0.33	0.33	0.31	0.34	**0.35**
Building	**0.33**	**0.33**	0.32	**0.33**	0.32
Bus	0.36	**0.60**	0.44	0.52	0.53
Dinosaur	**0.98**	0.95	0.92	0.95	0.96
Elephant	0.40	0.25	0.28	0.40	**0.45**
Flower	0.40	**0.63**	0.58	0.60	0.56
Horse	**0.72**	0.63	0.68	0.70	**0.72**
Mountain	0.34	0.25	0.32	**0.36**	0.35
Food	0.34	**0.49**	0.44	0.46	0.46
Average Precision	0.468	0.493	0.469	0.514	**0.519**

the system. The C, T, S order provides a higher performance.

This C, T, S order can be used with general datasets and if we need to apply this on a special dataset such as for objects, this order must be checked prior to use, as shape is the most important feature for object retrieval. Features are simply concatenate (simple feature combination) to generate feature vectors in these experiments and results could be improved by adding weights to the sub-features according to their single feature performance.

Figure 3.1 and figure 3.2 in Chapter 3 show some example images which covers Wang and Oliva and Torralba datasets. Each image represents an image category that we used for evaluation. It can be seen that there are visually similar images in different categories as mentioned early in this section.

4.3.6 Section Summary and Conclusions

This research work presents a simple yet effective novel image retrieval approach based on multi-level sequential searching using colour texture and shape features for the representation [Chathurani et al., 2015b]. This feature order is selected for general purpose datasets and this order can be changed according to the dataset. This approach is proposed to improve the retrieval quality of CBIR using global representation. The proposed system was evaluated and was compared to validate the proposed approach using two standard datasets. According to the experimental results, the proposed CBIR approach outperforms other existing systems approaches based on global descriptors on standard datasets in literature in terms of improvement in retrieval quality.

Parameters used for evaluation can be tested further to find optimal settings to achieve better retrieval performance. According to the results, this proposed method even outperforms some local representation techniques on the Wang dataset. This approach can be used not only for the global representation but for the local representation as well. Performance can be improved further by introducing appropriate feature weights.

However, this method has drawbacks as well. Feature dimension, feature interdependence and features values may have impact on retrieval performance. Therefore, Section 4.4 proposes a solution to latter problem. Still, the system may have to suffer from curse of dimensionality as it becomes the bottleneck for large datasets.

4.4 Image Retrieval Based on Late-Feature Fusion

In this research work, we describe a simple yet effective approach to achieve linear feature fusion with a combination of feature weights (significance) and distance normalisation (distance distribution), which can be applied to any combination of features. This general approach was invariant to the distance measures, dimensions and ranges of the features, as a pre-defined membership function was used. Empirical evaluation was performed on a subset of the standard Corel dataset to validate the performance of this proposed approach and it was compared against other implemented and independently evaluated approaches. This approach was further validated using Oliva and Torralba dataset [Chathurani et al., 2016b].

4.4.1 Background Work

A single image feature type is not adequate to differentiate images with the increasing size and variability of image databases. Therefore, to overcome the shortcomings of a single feature vector recognition algorithm, feature fusion such as a combination of colour, texture and shape features was introduced to CBIR [Hiremath and Pujari, 2007a, Mansoori et al., 2013] to cover a heterogeneous dataset. Multi-feature fusion is one of the ways to improve retrieval performance among other different techniques. The general solution for this technique can be obtained by combining two approaches:

I. feature engineering using distance combination, weightings and normalisation of features

II. using trained classifiers with the derived features to optimise performance with training data.

This work targets the improvement of the solution by focussing on feature engineering.

A simple feature fusion approach is to combine all the features to generate a single feature vector [Hiwale et al., 2015], or to obtain a summation of distances over different features [Hiremath and Pujari, 2007a]. But this simple approach assumes that all features carry equal importance. However, each feature has its own significance in image retrieval and in order to obtain effective outcome, the

varying degree of importance in each feature needs to be captured. Some systems achieve this by using methods such as weighting schemes [Yuan et al., 2011b,Saad et al., 2011, Ruikar and Kabade, 2016], different distance measures [Saad et al., 2011,Mansoori et al., 2013] or feature normalisation [Hiremath and Pujari, 2007a].

Feature fusion has a significant impact on CBIR and thus the performance of feature fusion is highly dependent on features, dimensions and ranges. As features have different variability, appropriately selected distance measures for each feature help to improve the retrieval performance. Feature normalisation maps a feature into a fixed range and feature distribution must be appropriately normalised.

There are different fusion techniques such as rank fusion and feature fusion. Five existing late-feature fusion methods, as shown in table 4.7 have been compared in [Chatzichristofis and Arampatzis, 2010] (where result-lists from individual descriptors are fused during query time). Each feature fusion method is brief in the table 4.7. It was found that the addition of all scores per image with normalisation (Z-score + CombSUM) outperform the other methods. Normalisation was done using Z-score (mean and standard deviation) in their method, which was different from the proposed distance normalisation approach which is described in detail in Section 4.4.5.

4.4.2 Image Features

Different low-level features were used, such as a combination of colour, texture and shape. Well-known Colour Histogram and Colour Moments, which extract different variations of colour, were selected as the colour feature to describe an image. YCbCr and CIEluv colour spaces were used, as those provide a closer match to human perception. Well-known Gabor Wavelet and Edge Histogram Descriptors were used as texture descriptors in this method. Invariant Moments were used for shape retrieval.

All these features were selected as they had shown good individual performances in the literature [Saad et al., 2011,Qiu, 2002,Rahmana et al., 2011,Agarwal et al., 2013] as well as being further validated through preliminary experimental evaluation. The performance of feature fusion does not depend on the individual performance of features, it depends on the diversification of the features as well as the inter-relation of features. Therefore, some feature combinations may degrade

Table 4.7: Some late-fusion methods compared in [Chatzichristofis and Arampatzis, 2010]

CombSUM	Addition of all scores per image, without any normalisation.
BC+ CombSUM	Borda Count [Aslam and Montague, 2001] originates from social theory in voting. Votes across ranked-lists are naturally combined with CombSUM.
Z-score+ CombSUM	Z-score is a linear normalisation per query which maps each score to its number of standard deviations above or below the mean score. Present that results with CombSUM.
IRP	The Inverse Rank Position [Jovic et al., 2006] merges ranked lists. It is the inverse of the sum of inverses of the feature similarity rank scores for each individual feature for a given image from relevant feature similarity ranking lists.
HIS+ multiplication	HIS [Arampatzis and Kamps, 2009] is a non-linear normalisation which maps each score to the probability of a historical query scoring a collection image below that score. Those probabilities combined with multiplication.

the retrieval quality more than the performance of the individual features when used in isolation.

A suitable combination of features had to be selected. We tested other features, such as Generic Fourier Descriptor, and Discrete Wavelet Transform using cross validation. However, experimental results were not promising with the combination of other features. We achieved the best performance with the combination of the five features described above from the separate experiments of sequential forward

selection (add one feature in) and sequential backward selection (take one feature out) of features. Mean Average Precision (MAP) is used as a performance measure.

The main feature selection criteria for the main system is described in Chapter 5

4.4.3 Feature Representation

Global image representation as well as local representation were used to validate the proposed late-feature fusion method. Results highlighted that the proposed feature fusion achieved better retrieval results. Features were extracted from the full image for the whole dataset for global representation. Grid representation was used as local representation. Firstly, the image was subdivided into nine non-overlapping blocks and then four overlapping blocks were generated by combining the sub-images by assuming that the main object of the image is generally located at the centre of the image. Five individual feature vectors were used to represent each image. The performance of the proposed fusion technique was not affected by the range and the length of feature vectors. So it maintained the normal form with absolute values.

4.4.4 Weights Calculation

Weight assignment for features is important in multiple feature fusion as different features have different significance. Each image in the database could be represented as follows;

$$F_I = [f_1, \quad f_2, \quad f_3, \quad f_4, \quad f_5]$$

Feature Index	f_1	f_2	f_3	f_4	f_5
Feature Name	ch	cm	gabor	ehd	im

w_i is the weight related to i^{th} feature f_i and w_i was considered as 1 (same significance) for each feature in simple feature fusion. Different weights were used in weighted feature fusion according to their relative single feature performance ($W_{f_1} = 0.266$, $W_{f_2} = 0.159$, $W_{f_3} = 0.218$, $W_{f_4} = 0.233$, $W_{f_5} = 0.124$) where

$$\sum_{i=1}^{5} W_{f_i} = 1$$

.

Precision was used as the performance measure which is described in Chapter 3. Weight values were calculated according to the MAP values. MAP was calculated for each feature using whole database as queries. The higher the MAP for a feature, the better the ability to retrieve correct images, the higher the weight related to it. This is a general solution for weight calculation. If we have a well-categorised specific dataset to improve results further, we can assign different weights for different categories for one feature (by considering inter-class variation), but that solution will be specific to the selected dataset.

4.4.5 Membership Function

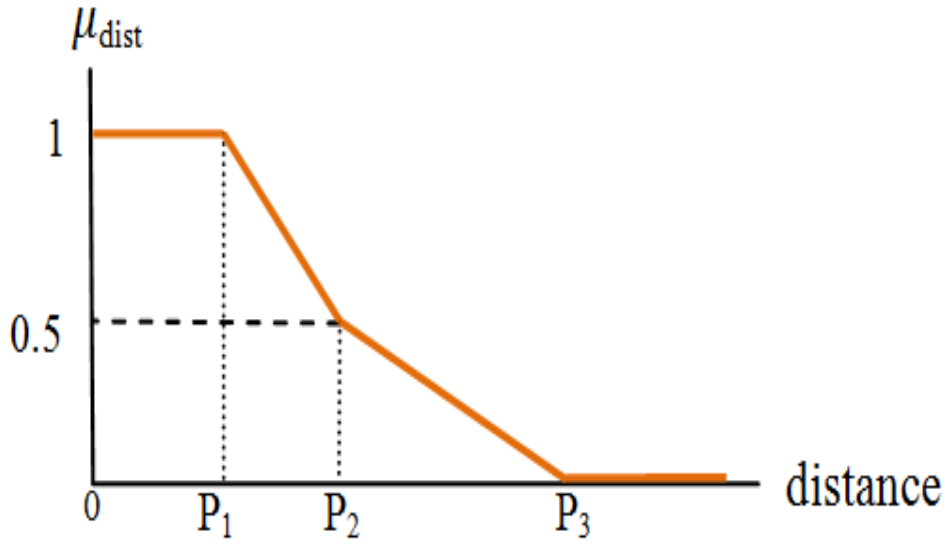


Figure 4.4: Distance regions generated by piecewise linear function.

A simple piecewise linear function was used to generate rules. It was easily implemented and all the values were mapped to the interval [0, 1]. Regions were defined for each feature according to the distance measure in this function as shown in figure 4.4 ($0 - P_1$, $P_1 - P_2$, $P_2 - P_3$, $P_3 <$). Four regions were selected by defining three points according to the ranked distance in ascending order as best (first), average (middle) and worst (last). We defined regions according to the image similarity and most similar, least similar and averagely similar to the query was used. The average distance of the first five images, middle five images and last

five were calculated individually from the listed n number of images which were used as a training set, and an average was calculated (This n will be described later in this section). Then these calculated averages of all the categories were used to calculate point P_1, P_2, P_3 respectively for each feature. As an example, the calculation of P_1 for feature f_i (P_{1_i}) is shown in equation 4.7:

$$P_{1\_i} = \frac{\frac{1}{N}\sum_{n=1}^{N} dist_{n\_f_i}}{C} \tag{4.7}$$

Where N = number of images considered (N=5), C = number of categories in the data set (C=10 and C=8 for Wang and Oliva and Torralba datasets), $dist_{n\_f_i}$ is the n^{th} ranked distance related to the feature f_i.

When searching images, a membership value must be computed for each feature vector by using the calculated distance. Equation 4.8 was used to map the distance to the value in the range of [0,1] (least similar to most similar). A random n number of images were taken out from each class (half of each class was used as a training set i.e. n=50 for Wang dataset) to generate this distance membership function and there were 500 and 1344 images altogether in the training set for the Wang and Oliva and Torralba datasets respectively. This n number of images were selected randomly and the training set was changed from time to time by selecting different image sets to confirm that the performance of the proposed approach did not vary with the selected training set, which meant that performance was not heavily dependent on the selected dataset and not optimised for a particular training set. Finally, the average was taken into consideration.

$$\mu_{dist\_f_i} = \begin{cases} 1, & \text{if} \quad dist_i \leq P_{1\_i} \\ \left[\frac{0.5}{P_{1\_i} - P_{2\_i}}\right] dist_i + \frac{1 - 0.5 * P_{1\_i}}{P_{1\_i} - P_{2\_i}}, & \text{if } P_{1\_i} < dist_i \leq P_{2\_i} \\ \left[\frac{0.5}{P_{2\_i} - P_{3\_i}}\right] dist_i + \frac{0.5 - 0.5 * P_{2\_i}}{P_{2\_i} - P_{3\_i}}, & \text{if } P_{2\_i} < dist_i \leq P_{3\_i} \\ 0, & dist_i > P_{3\_i} \end{cases} \tag{4.8}$$

4.4.6 Similarity Measure

The performance of a CBIR system mainly depends on the particular image representation and similarity matching function employed. Colour, texture and shape features were extracted for this image representation, as we were targeting general images. Related feature weights and membership values related to the distance of the query were used in the similarity measure. The proposed approach is simple and easy to adapt and it is described below.

Similarity was calculated by using defined weights and membership values for each feature in the proposed approach. Different features have different significance and hence, the significance of each feature was considered for multi-features-based retrieval. Euclidean distance was used as the distance measure. Similarity between image Q and I can be calculated as below:

Step 01: First, the distance is measured between image Q and image I for feature f_i ,$Dist(f_{iQ}, f_{iI})$ and computed $dist_{f_i\_QI}$.

Step 02: Membership value $\mu_{dist\_f_i}$ is computed for feature f_i from the distance membership function by using the $dist_{f_i\_QI}$ in equation 4.8.

Step 03: The weight of the feature f_i is considered as w_i.

Step 04: Repeat the steps 1 to 3 for each feature f_i (we use 5) that is in the system and computed the membership value $\mu_{dist\_f_i}$ and weight value w_i.

Step 05: Similarity measure $Sim(Q, I)$ is computed, fusing all the feature measures using equation 4.9.

$$Sim(Q, I) = \sum_{i=1}^{N} w_i * \mu_{dist\_f_i} \tag{4.9}$$

Where N is number of features.

Step 06: Repeat the steps 1 to 5 for whole dataset and list them all. Then rank the list according to $Sim(Q, I)$ and retrieve the top-ranked images.

4.4.7 Experimental Results

To evaluate the effectiveness of the proposed feature fusion approach, experiments were performed on two general purpose image datasets; Wang and Oliva and Torralba. The details of these datasets can be found in Section 3.1.1 and Section 3.1.2 in Chapter 3.

Experimental setup

The performances of global individual features were considered to calculate relative weights and the membership function. Since the goal of feature fusion was to achieve better retrieval results than any single feature, the best result of single feature (global ch) performances was used as the baseline.

The most common evaluation measure in information retrieval is precision and it is calculated using equation 3.2. Fusion-based similarity measures were compared based on average precision by evaluating the top 20 retrieval results. A retrieved image was considered a correct match if and only if it was in the same category as the query image.

Moreover, Precision@n was calculated using equation 3.3 and equation 3.4 for n=5 to 50.

In these experiments 1000 and 2688 images were used, and half of the images were used for training and other half were used for testing. It may be noted that this method solves the problem of high dependency on feature dimensions and ranges in feature fusion. However, training is essential each time this is applied to a new database.

Results

Figure 4.5 shows the performance comparison of feature fusion with the baseline for each class (AP), where performance of the colour histogram was considered as the baseline. The different feature fusion methods given bellow were compared. These experiments were carried out to study the improvement over simple feature fusion, weighted feature fusion to the proposed feature fusion (weights + distance normalisation).

i. Simple global and local feature fusion (concatenation) by considering each feature with equal significance for retrieval.

ii. Weighted global feature fusion, weighted local feature fusion.

iii. Global feature fusion with weights and distance normalisation, local feature fusion with weights and distance normalisation.

Figure 4.6 further elaborates the performance. Local representation (grid) showed higher performance (MAP=0.67, MAP=0.7, MAP=0.72, for case i, ii and iii respectively) compared with the performance of global representation (MAP=0.62, MAP=0.63, MAP=0.66, for case i, ii and iii respectively). Weighted feature fusion showed higher performance (MAP=0.7 for local and MAP=0.63 for global) than the performance of simple feature concatenation (MAP=0.67 for local and MAP=0.62 for global). Performance (MAP=0.72 for local and MAP=0.66 for global) of weighted feature fusion combined with distance normalisation gave the highest performance in both global and local representation. According to the results obtained, the proposed approach shows a higher performance than weighted and simple feature fusion.

Figure 4.7 shows the performance comparison with other systems which had been used the Wang dataset for evaluation. Our system outperformed the other systems by obtaining MAP of 0.72. All the other systems (the addition of all scores or merging features [Hiremath and Pujari, 2007a, Yuan et al., 2011b], weighted distance [Saad et al., 2011, Mansoori et al., 2013]) showed MAP less than 0.66 except one which was proposed in [Chatzichristofis and Arampatzis, 2010]. In [Chatzichristofis and Arampatzis, 2010] authors had tested five feature fusion methods as shown in table 4.7 and found that the addition of all scores per image with normalisation (Z-score + CombSum) achieves the best performance. Please refer [Chatzichristofis and Arampatzis, 2010] for detailed description of these methods as we considered only the best performed one from [Chatzichristofis and Arampatzis, 2010]. Z-score + CombSum method was tested on Wang dataset and it achieved only MAP of 0.67 for local feature fusion (second best performance of compared performances). Z-score + CombSum was the best among five feature fusion approaches that were compared and our proposed approach was superior to that best late-fusion method described in [Chatzichristofis and Arampatzis, 2010]. The proposed approach showed superior performance in both local and global representation.

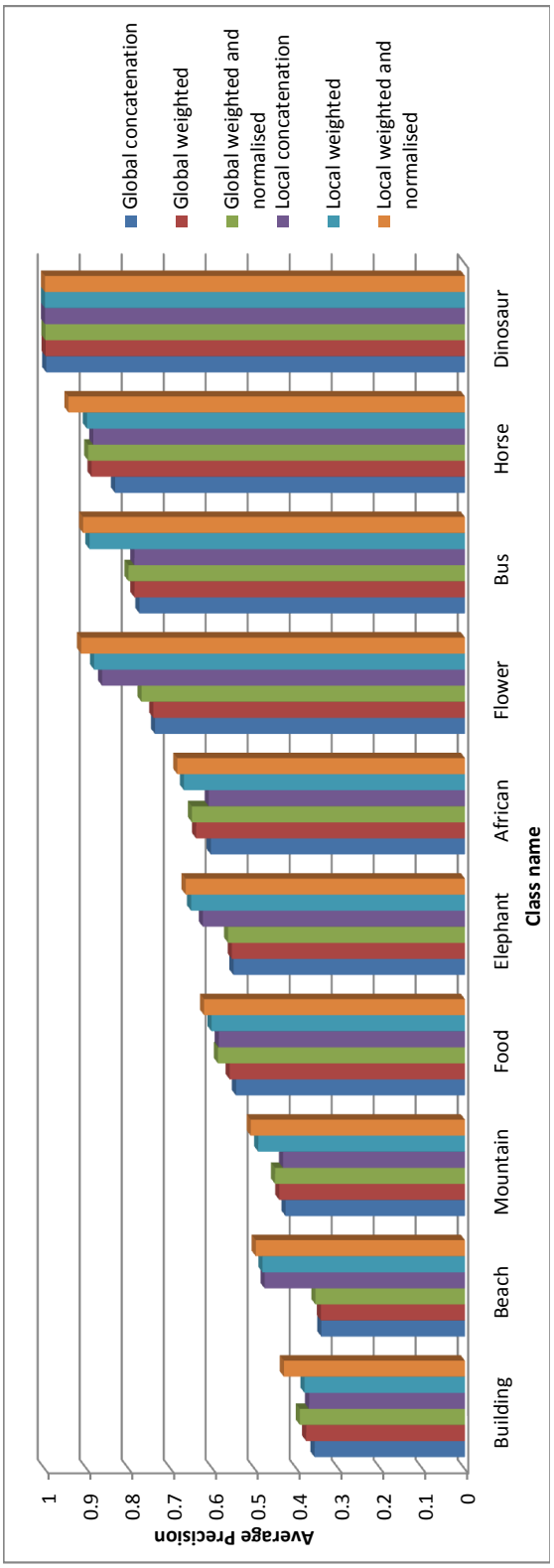


Figure 4.5: Performance comparison of linear feature fusion for each class in Wang dataset.

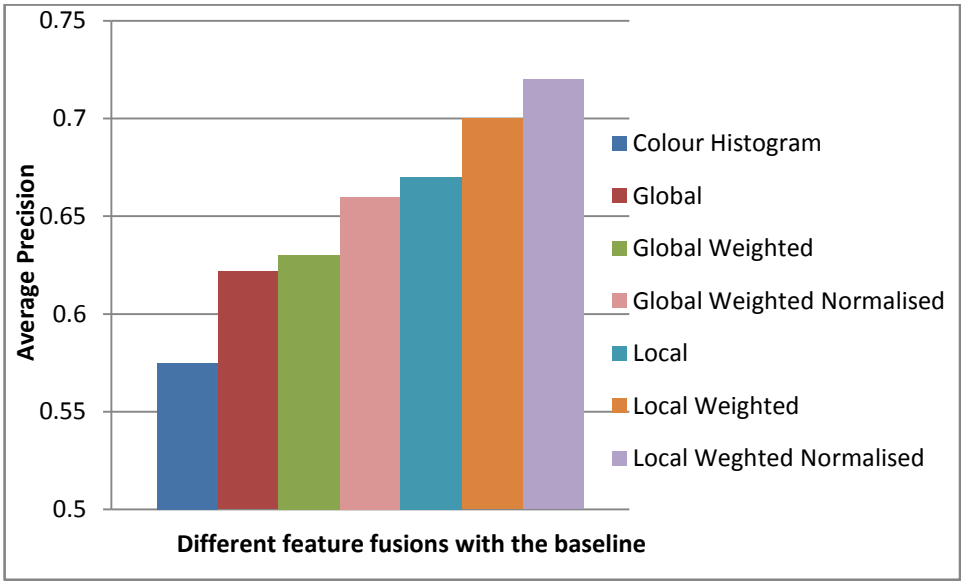


Figure 4.6: Performance comparison of feature fusion with the baseline on Wang dataset (AP@20).

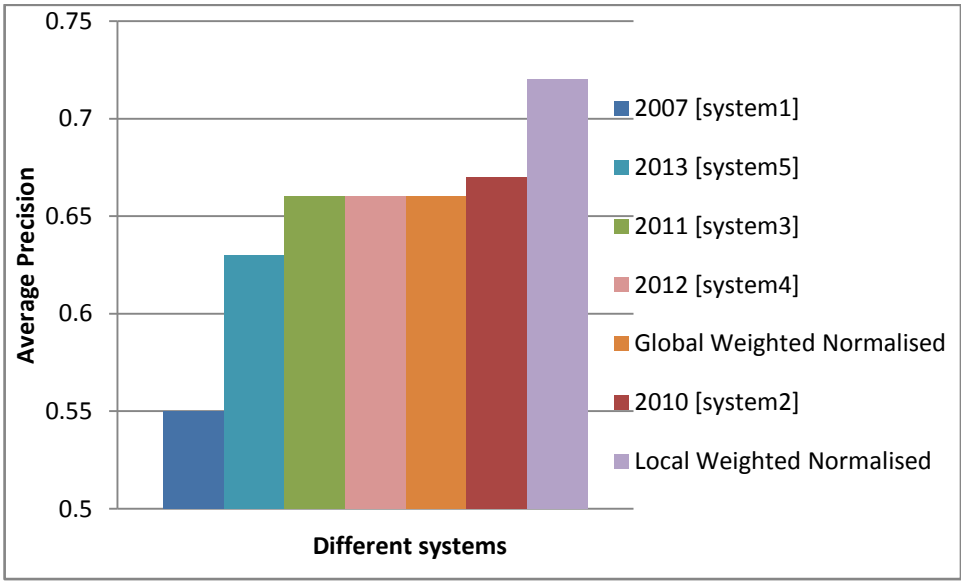


Figure 4.7: Performance comparison of different systems on Wang dataset (AP@20). System2 [Chatzichristofis and Arampatzis, 2010] (Z-score + CombSum) is the best late-fusion method from the compared methods in table 4.7. (system1- [Hiremath and Pujari, 2007a], system5- [Mansoori et al., 2013], system3- [Yuan et al., 2011b], system4- [Saad et al., 2011])

The proposed approach [Chathurani et al., 2016b] was further validated using the Oliva and Torralba general purpose image dataset. Figure 4.8 shows that the performance comparison of different fusions are the same as in figure 4.6. The proposed feature fusion method showed an improvement in performance on this dataset as well, seeing that our method achieves 0.63 MAP while Z-score + CombSum achieves 0.59 MAP. So it was found that the proposed approach can be applied to any database and it is not optimised for one dataset. While this approach has its advantages as mentioned, the main drawback of the approach is to be trained in the beginning which is an off-line process.

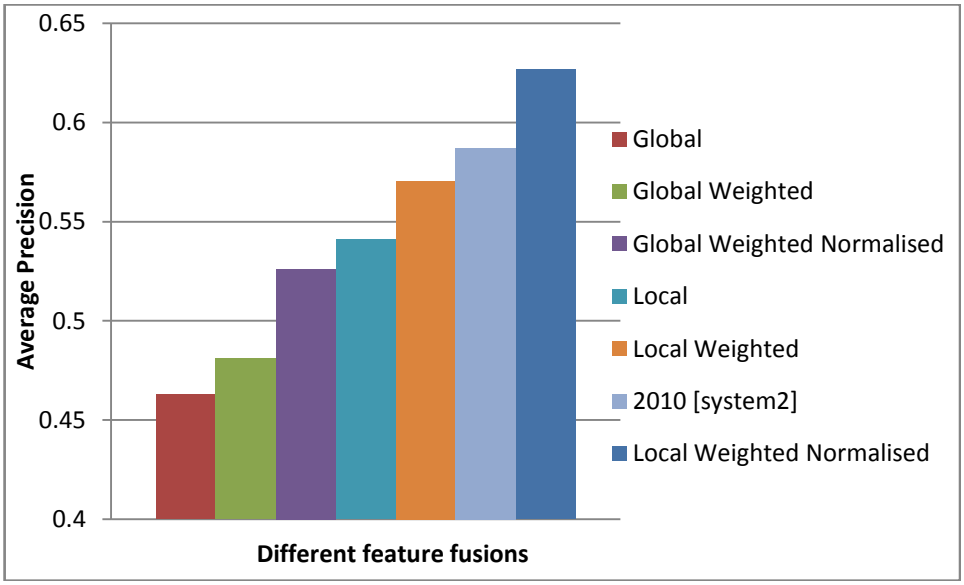


Figure 4.8: Performance comparison of feature fusion for Oliva and Torralba dataset with Z-score + CombSum fusion (AP@20). (system2- [Chatzichristofis and Arampatzis, 2010])

Retrieval quality of the proposed approach was assessed by calculating MAP@n (n is the number of images retrieved at a time) on both the datasets. Figure 4.9 shows MAP @ n for Z-score + CombSum [Chatzichristofis and Arampatzis, 2010] and our method (weight + normalisation). Here it is shown that our late-fusion method is better than the other methods and it has good retrieval performance.

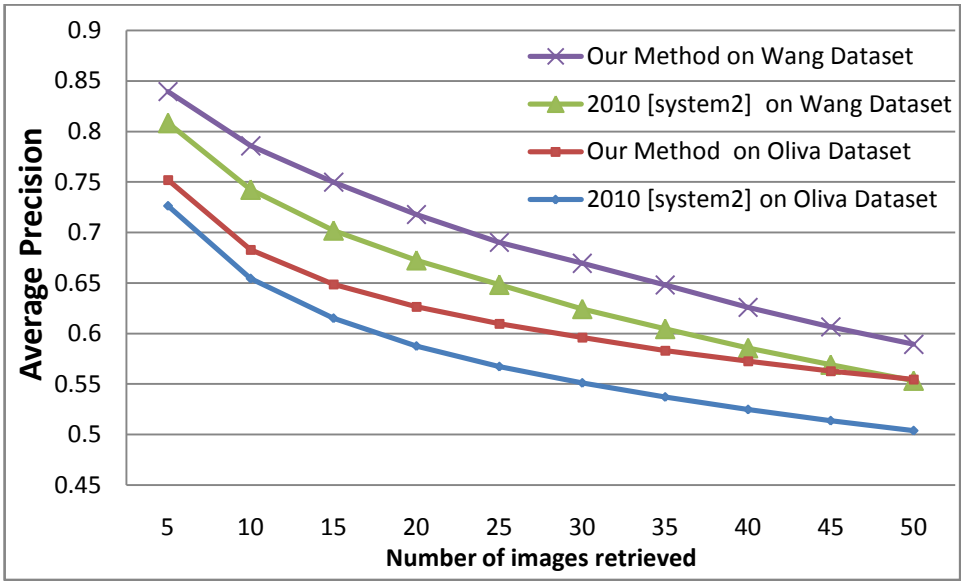


Figure 4.9: Mean Average Precision at N (MAP@N). (system2- [Chatzichristofis and Arampatzis, 2010])

4.4.8 Section Summary and Conclusions

First, the best feature combination was found from different low-level features using cross validation. Then using those features, a simple fusion-based similarity matching approach was proposed based on a weighted combination of similarity measures of different features according to their relative performance and distance normalisation. A simple membership function was used to normalise the distances to the [0,1] interval to remove the effect of biasing due to the length of the feature vectors and the large values of distance. The proposed approach can be easily adapted in any feature combination. The proposed approach was tested on global representation as well as local representation and we observed improvement in retrieval quality. Moreover, the proposed system showed superior performance in retrieval quality relative to the existing feature fusion approaches. Dynamic feature weighting can be used according to the given query image to improve retrieval quality further as a modification.

However, this is not scalable for large databases as high dimensonality of feature vector leads to curse of dimensionality and training is required to apply on a new dataset.

4.5 Chapter Summary and Conclusions

In the beginning, different image descriptors were selected according to their retrieval performance and several experiments were carried out with those features as preliminary work. Two basic image retrieval techniques were proposed based on feature normalisation and multi-level searching. The results of the empirical evaluation of these systems on different datasets concluded that these methods are useful for CBIR. But those proposed methods had drawbacks, such as scalability issues for large databases due to computational complexity and training was required. Therefore, these methods were not extended in this thesis and we studied how we can address these issues and then proposed a signature-based solution which is introduced in Chapter 5.

However, from here we learned the suitable features which can be used in the proposed CBIR as demonstrated in Chapter 5. The image features selected for use in this research after these studies and the feature selection criterion will be discussed in Chapter 5, as the outcome of that was benefited by the existing system, which is described in Chapter 5.

Chapter 5

Image Signature Representation

Chapter Organisation

This chapter presents the development of a content-based image retrieval (CBIR) system using image signatures, namely CBIR-ISIG. Section 5.1 provides an introduction to CBIR and the features it uses. The steps required to follow to design and implement the signature-based CBIR is presented in Section 5.2. Indexing of images signatures is explained in Section 5.3, followed by the performance evaluation analysis of results in Section 5.4. Application of the signature-based image retrieval is presented in Section 5.5. The chapter summary and conclusions are included in Section 5.6. The original contributions discussed in this chapter resulted in publication (i), (iii) and (vii) in List of Publications. The image signature representation presents in this chapter by using the features studied in Chapter 4 is the core of the systems considered in this thesis, and in the next chapters it will be used as baseline.

5.1 Introduction

Developing a CBIR system for general-purpose image databases with semantically accurate retrieval is cumbersome due to a number of reasons, such as the large size of the database, the gap between the high-level semantic concepts required by the system users and the low-level visual features extracted from the images, namely semantic gap. Many researchers have addressed this problem from different points

of view. Local representation is a simple approach that tries to achieve better retrieval performance by reducing the semantic gap. Local representation can be derived from the points of interest, sub-images or regions which are gained by decomposing the full image into parts. In the literature, most region-based approaches rely on image segmentation. However, a precise image segmentation method that can be applied to general image collections has not yet been found. In the absence of an accurate segmentation approach, a sliding window approach over location and scale has shown to be quite effective in [Hiremath and Pujari, 2007a]. Therefore, grid-based CBIR approaches have been proposed [Takala et al., 2005, Hiremath and Pujari, 2007a, Smeulders et al., 2000, Rahman et al., 2006] to improve retrieval performance, but only a few studies have been done in the past. The primary advantage of the grid-based approach over segmentation is less computational complexity.

Therefore, grid-based local representation was selected to be used in the system. Image representation was derived through sub-image decomposition and BoW approach was adapted. Initially, this work focuses on identifying various sizes of semantic image feature building blocks which can be used to represent an image as a bag of semantic image features. The main building block of this research was image block which was derived from grid-based image decomposition. Image features were extracted from selected image blocks and independent visual vocabularies were generated from each feature using K-means clustering.

Feature Extraction is an important step and it is responsible for the quality of the retrieval performance in CBIR. Colour, texture and shape features have been used in CBIR systems in different ways [Hiremath and Pujari, 2007b, Saad et al., 2011, Mansoori et al., 2013, Li and Wang, 2003, Hiremath and Pujari, 2008, Chowdhury et al., 2012]. Colour is an important feature and there are number of colour spaces available for use. Texture feature is important when describing the real world images. Shape feature helps when dealing with objects. General images may contain all the types of images, therefore, a combination of these features [Hiremath and Pujari, 2007b, Yuan et al., 2011b, Saad et al., 2011, Li and Wang, 2003, Li et al., 2000, Chen and Wang, 2002, Hiremath and Pujari, 2008] is the best solution for better performance. As a single feature is not adequate to differentiate images, a set of descriptors are used to extract image features

which covers colour, texture and shape features. The initial features used for the experiment are described in Chapter 4. Finally, the best feature combination was selected to use in the proposed CBIR system namely CBIR-ISIG, and the feature selection and selected features are described in this chapter.

Among different techniques for local feature representations, BoW is one of the most powerful techniques and has shown good retrieval performance [Yuan et al., 2011b, Mansoori et al., 2013]. In the BoW approach, visual vocabulary was performed by grouping similar local features together. Each of these groups has a centre, and that cluster centre was treated as a word. Finally, a histogram was generated for each image by mapping sub-image features to the cluster centre. Typical BoW approaches have used histograms by frequency of the occurrence of each word, with different variations. Instead of generating a typical histogram of visual words for each image, an image was converted to a symbolical representation.

The descriptors' dimensionality highly influences the performance of CBIR. Therefore, different dimension reduction techniques have been proposed and used in the literature, such as latent semantic analysis and principal component analysis [Gorman and Curran, 2006, Elharar et al., 2007, Banda et al., 2013]. However, these are computationally complex and relearning is required with a new addition, which is time consuming. Random indexing (RI) has been used and has shown great promise as a dimensionality-reduction technique in text retrieval [Vries et al., 2009, Gorman and Curran, 2006]. Compared with the other methodologies, RI has low computational cost, lower complexity, competitive accuracy, and most importantly, it is an incremental approach. RI can be used in image retrieval and can achieve all the benefits of it. If we consider the BoW approach, it is a precise representation but it suffers from the curse of dimensionality. Therefore, RI can be used to reduce the feature space of BoW representation. Finally, image signatures are generated from BoW representation and those image signatures are fixed length binary strings. Those binary signature have a control over retrieval speed by reducing the feature space.

The performance of the proposed approach was evaluated using three benchmark datasets for the retrieval quality which highlighted that the proposed approach has a high potential to retrieve correct images. System performance was compared with existing systems in the literature and the results indicated that our

approach has superior performance over the other systems.

Finally, application of this signature-based image retrieval was introduced by defining the rotation invariant BoW approach using the signature approach, especially for object detection. This approach compared with typical BoW approaches and the results outperformed those methods on object datasets.

5.2 Image Representation

This research uses both global and local representation in different cases. Initially global representation was used as mentioned in the Chapter 4 but mainly, local representation was used. This research used grid-based representation as local representation. Moreover, a circular image decomposition method was proposed.

5.2.1 Features for the System

General images may contain all the types of images and therefore need a combination of colour, texture and shape features to address a general collection. Therefore, we selected descriptors covering all the features. However, there must be a good combination of features and the most suitable features must be selected to describe images in the system. Therefore, we selected features using an off-line feature selection algorithm. We used leave-one-out and add-one-in for the feature selection and realised that both gave same results. Therefore, only leave-one-out experiments will be described in this section. In this method, initially image retrieval performance was evaluated using a full feature set and then determined if a feature was bad by comparing initial performance with the image retrieval performance by leaving a feature out for every feature. Then the worst one was left out and the process was repeated. A stopping criterion determines when the feature selection process should stop and the stopping criterion used in this research was the subsequent deletion of any feature that did not produce a better subset (no significant change or sign of it starting to deteriorate the performance when removing). Finally, all the features which deteriorate the performance were eliminated and the remaining features were selected as the feature set to use in the system. Initially, Colour Histogram (CH), Colour Coherence Vector (CCV), Colour Moment (CM), Discrete Wavelet Transform (DWT), Gabor Wavelets (GABOR), Edge Histogram Descriptor (EHD), Image Moment (IM), Generic Fourier Descriptor (GFD) and

GIST features were selected and experiments were carried out using these features and seven features were selected for the proposed signature-based CBIR.

Several iterations were carried out and some significant results are shown in the figure 5.1, figure 5.2 and figure 5.3 with a brief explanation. Figure 5.1 shows the retrieval performance variation when one feature is left. According to that, there are four bad features and GFD has the highest positive change which means it is the worst feature and that feature was removed. Figure 5.2 shows the retrieval performance variation when one feature is left with respect to the results by eliminating GFD. According to that, there are four bad features and DWT has the highest positive change which means it is the worst feature and that feature was removed. Figure 5.3 shows the retrieval performance variation when one feature is left with respect to the results by eliminating GFD and DWT. According to that, results are deteriorating when other features are removed.

Experiments were carried out by leaving all the bad features and we found the subset of features from the full set by eliminating DWT and GFD. Therefore, seven features were used in this research work.

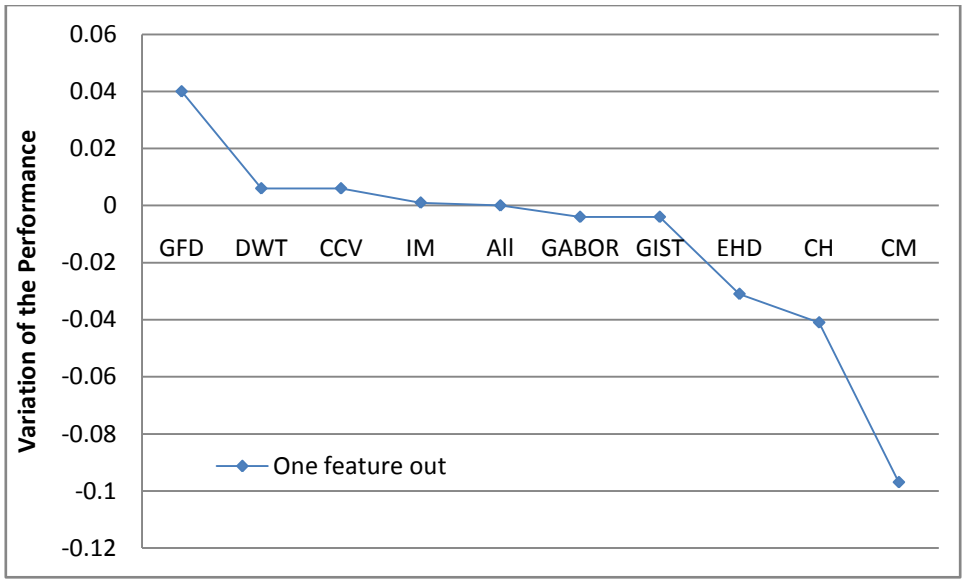


Figure 5.1: Leave one feature out - performance variation with reference to the performance with the full feature set.

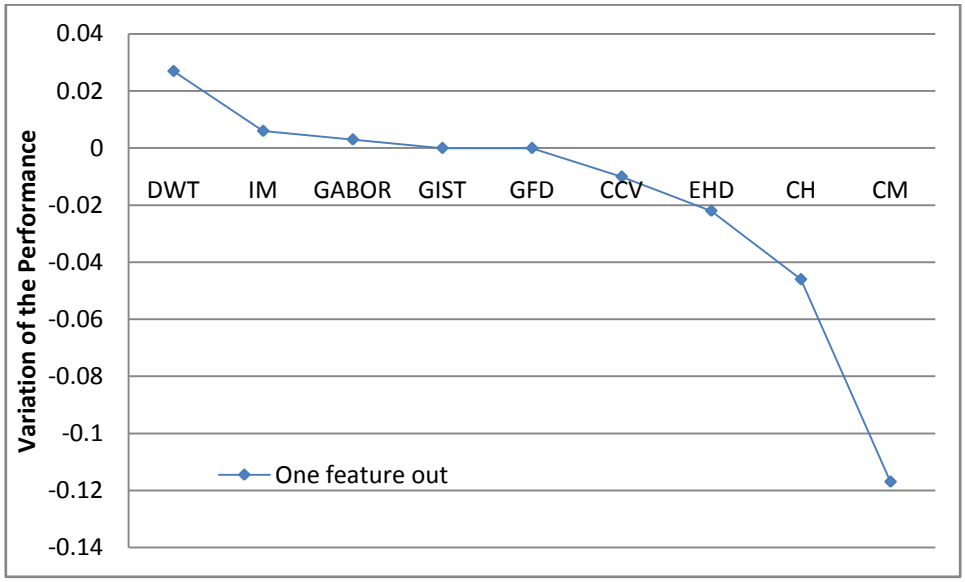


Figure 5.2: Leave one feature out- performance variation with reference to the performance of feature set without GFD.

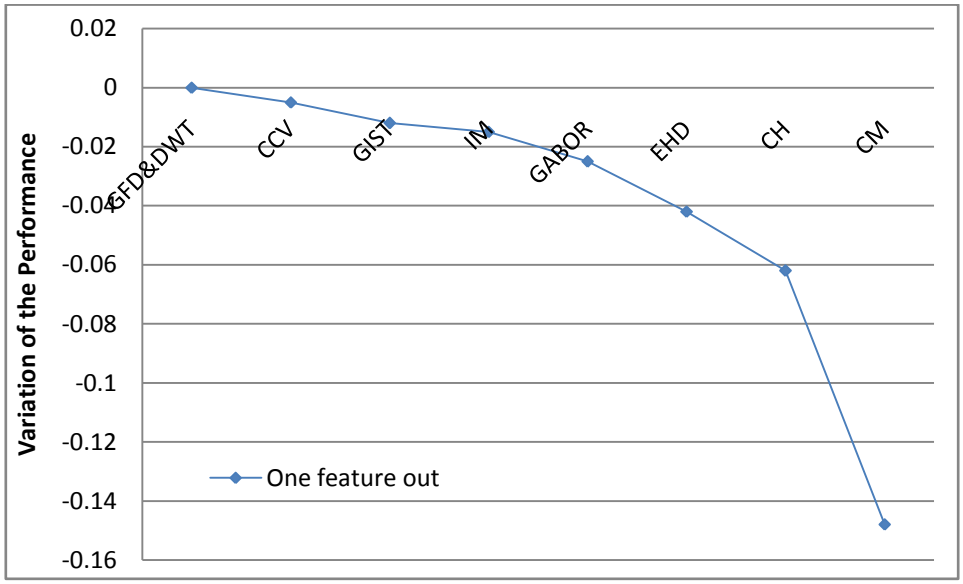


Figure 5.3: Leave one feature out - performance variation with reference to the performance of feature set without GFD and DWT.

5.2.2 Image Decomposition

Firstly, an appropriate method to decompose the image was devised in this research on grid-based implementation. Different image decomposition methods were tested and from that experimental evaluation it was found that a 3 by 3 grid framework is the best way to partition the images for BoW representation. Each image was divided into non-overlapping sub-images, as in figure 5.4.a and figure 5.4.b, and generate 9 sub-images.

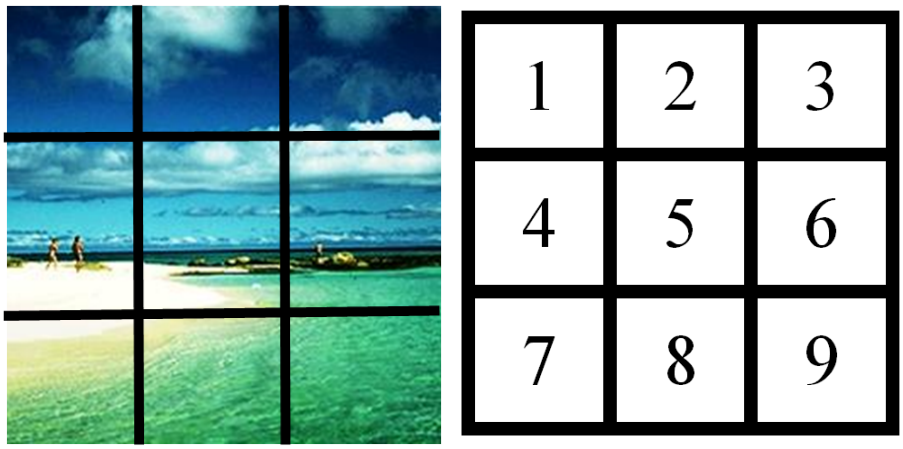


Figure 5.4: Sub-image generation using grid-based approach (a) and (b).

However, this research proposed a new circular image decomposition method, having the hypothesis that the circular image decomposition will improve the performance than grid-based method as this representation can achieve scale invariance property with the help of normalisation as image division starts from the centre of the image, if the object stays in the middle. Those sub-images can touch a bigger area than a 3 by 3 grid, and all the sub-images will touch the object, if the object stays in the middle of the images. Circular image decomposition is as shown in figure 5.5.

These were the main blocks which were used to extract features from images. The proposed image decomposition methods were tested on three benchmark datasets and it is described in the evaluation section (Section 5.4).

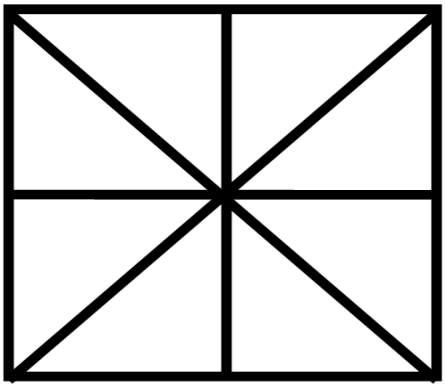


Figure 5.5: Circular decomposition approach.

5.2.3 Bag of Words Representation

Extracted image features must be represented in a way that the proposed method can be used to search images. As features were extracted from the sub-images, the feature database was considerably larger than image database as one image was represented by several sub-images and each sub-image was represented by several features. Therefore, precise representation was required to select which could be used to convert an image into symbolic representation (literally considers images as text documents). Therefore, it was necessary to adapt the BoW approach which can generate visual words from the extracted image features (real-valued vectors), and represents each image using set of visual words (represented as symbolic tokens).

It was necessary to select an appropriate multidimensional indexing algorithm to index the features. Clustering is a promising technique among indexing techniques. It is necessary to cluster the image features in order to obtain discrete representation of feature sets. K-means is one of the simplest and best-known unsupervised clustering algorithms that can be easily implemented for feature vocabulary generation. Therefore, K-means was used for clustering.

The process of converting the images to the symbolic representation is described in steps as follows:

Step 01: Firstly, features were extracted from the sub-images. Each feature was given an index f_i to denote it within the symbolic representation.

where $f_i \epsilon \{1, ..., N\}$ and N was the number of features used in the system.

Step 02: Visual vocabulary for each feature was generated independently using feature sets using simple K-means. Each feature set was clustered and each centroid was given a cluster number c_k where $c_k \epsilon \{1, ..., C\}$. Here, C was the vocabulary size. Each centroid was considered as a word.

Step 03: Appropriate visual words were found from these vocabularies through codebook lookup of each raw sub-image feature.

Step 04: Each sub-image was represented using words as

$$I = f_i\_c_k \qquad i\epsilon 1, ..., N \quad and \quad k\epsilon 1, ..., C \tag{5.1}$$

where f_i was the feature index and c_k was the index of the nearest cluster centre.

Step 05: Finally, the full image was represented as $IM_{full}$where

$$IM\_full = I\_a \qquad a\epsilon 1, ..., M \tag{5.2}$$

Here, I_a was a sub-image representation and M was the number of sub-images in an image.

This BoW representation has a different representation than a typical histogram based representation which counts the occurrence of each word appearing in the image. The images were represented symbolically, just like text, by using the codebook label of each cluster as a visual word to encode the feature as described in step 5.1 and 5.2. Then the images were converted into symbolical representation. Sub-images could be considered as paragraphs in a document, thus the full image could be considered as a document.

K-tree can be used for very large and dynamic datasets as it offers excellent run time performances and the dynamic properties of the tree [Geva, 2000, Vries et al., 2009, De Vries and Geva, 2009b].

5.2.4 Signature Representation

Image signatures were generated for sub-images as well as the full image. This was done to significantly reduce the dimensionality of the representation. The descriptors' dimensionality is important and heavily influences the complexity of the similarity measure in retrieval, and the memory requirements for storing the descriptors. Therefore, it is important to consider the feature dimension to improve searching speed. The BoW approach provides precise representation but still deals with high-dimensionality data, which presents scalability challenges. Therefore, the dimensionality of the representation is reduced as much as possible while keeping enough information to differentiate images.

There are several approaches available for dimensionality reduction including RI [Sahlgren, 2005], which is an efficient, scalable and incremental approach based on random projection to avoid the computational cost for matrix factorisation [Geva and De Vries, 2011]. One prime advantage of RI is that it can work directly with symbolic features. For instance, RI works directly with words in the converted symbolic representation. Therefore, RI is used effectively in text retrieval applications to reduce the dimensionality of documents without significant degradation in retrieval quality. Furthermore, it can produce binary image signatures. The representation of objects as bit vectors lends itself to efficient processing, with low level bitwise operations supported on all conventional processor architectures. Most importantly, RI can be performed incrementally, aligning with the new data arrival and it is crucial for online systems. Therefore, in this research, the RI approach was used for dimensionality reduction and to create image signatures. This allowed the feature vector space to be reduced in dimensionality without expensive factorisation such as latent semantic analysis (LSA) techniques. Seeding a pseudo-random number generator with the feature hash and then generating a feature signature. That was used to create a pseudo-random sparse ternary feature vector having values from { -1 ,0, 1 }. A common choice with RI was to assign the proportion of vector elements with each value { -1, 0, +1} to be $1/6, 2/3$ and $1/6$ respectively. All feature vectors of the entire image and its sub-images were then summed up to produce a single image index vector. The image index vector was then squashed into a binary signature by assigning bit '1' to positive values

and bit '0' to negative values. Similar images sharing similar features contained similar signatures.

Image signatures could then be compared for similarity by taking the bitwise (Hamming) distance between them. This technique can be used as a highly efficient replacement for a cosine similarity calculation in the original feature vector space. This research used a signature search-engine for searching. The motivation for using signatures to represent images comes from the fact that computation time quickly becomes a bottleneck when dealing with large databases and signature search engines can retrieve results from web-scale collections in milliseconds [Chappell et al., 2013]. Topsig [Geva and De Vries, 2011], which is available in open source, was used to generate and search signatures in our CBIR system. This research work was concerned with identifying the ability of our approach to represent and then find images, rather than the signature matching and searching mechanism itself, so in fact, it allowed us to use any signature search engine regardless of specifications. Our concern was with how well the signatures would represent the images.

5.3 Indexing

As this research used image signatures as the final image representation, a special indexing mechanism was not used. So image signatures were kept as a list. Signature search engines search millions of images in a sequential search. This approach used a signature search-engine to search image signatures [Geva and De Vries, 2011, Chappell et al., 2013] which is available in open source. Therefore, this approach inherited the scalability of the signature search engine. As image signatures were in binary format, signatures were compared by taking the bitwise (Hamming) distance between them for searching. This technique can be used as a highly efficient replacement for a cosine similarity calculation in the original feature vector space [Geva and De Vries, 2011].

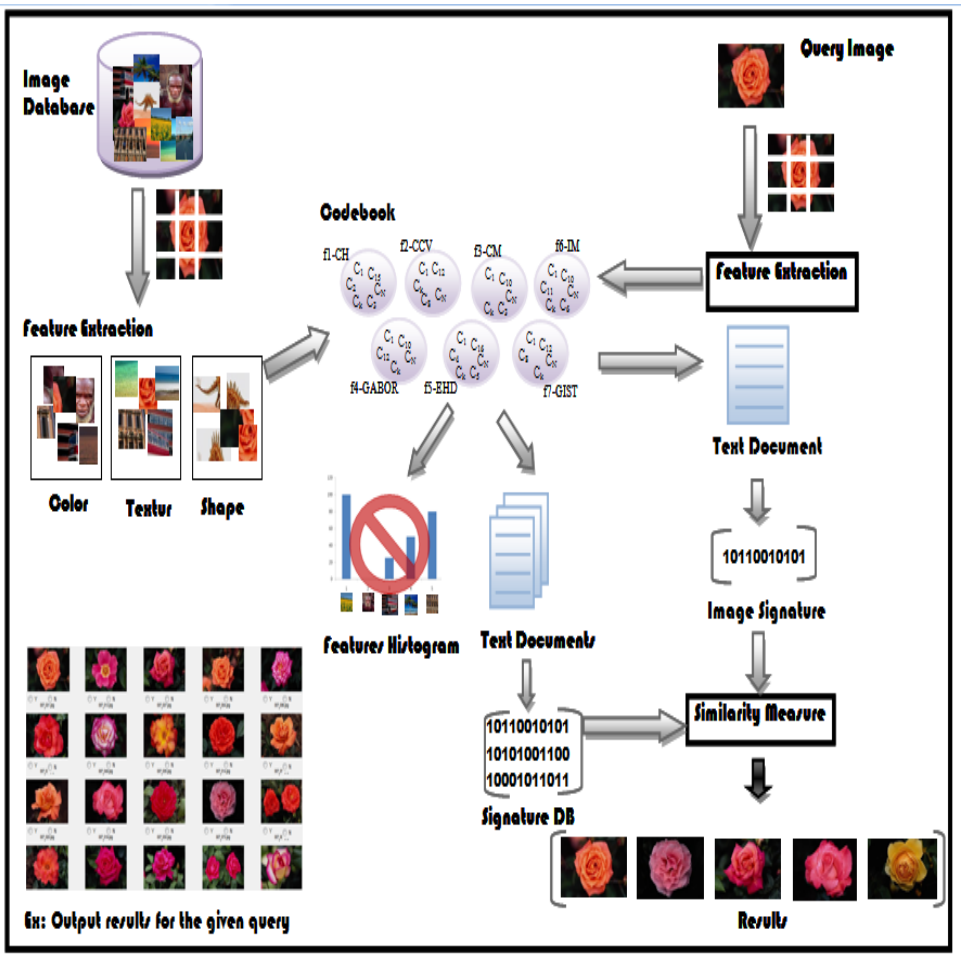
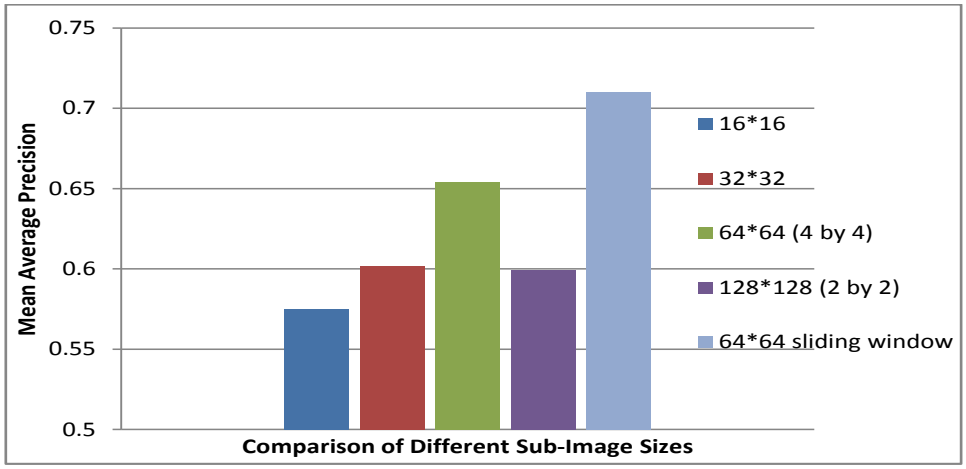


Figure 5.6: The Framework of Signature Generation and Searching in CBIR-ISIG system.

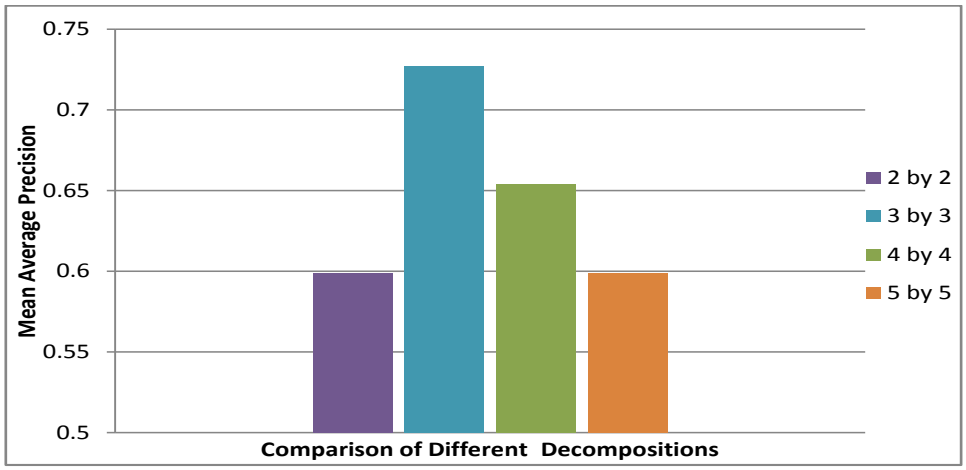
5.4 Evaluation

5.4.1 Evaluation Measures

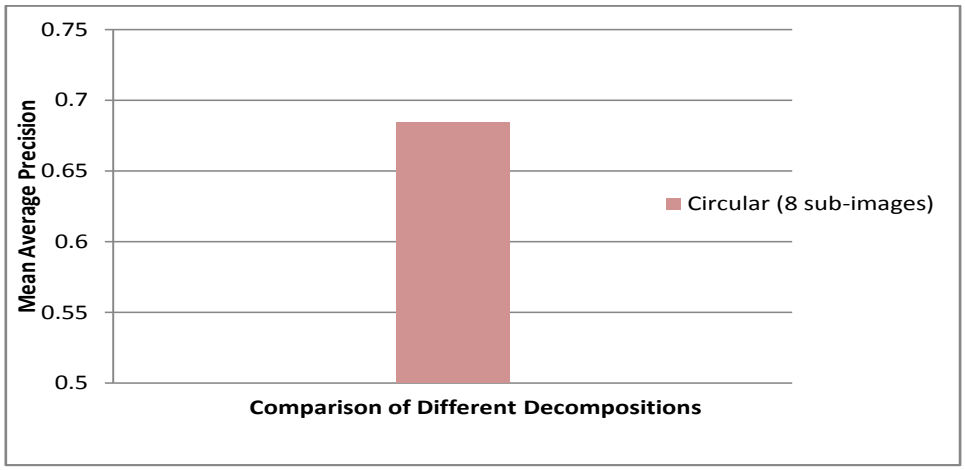
The CBIR-ISIG system was evaluated on different datasets using different evaluation measures to study the effectiveness of the approach in CBIR. Experiments were performed on several general purpose image datasets Wang, Oliva and Torralba and Flickr25K. Detail of these datasets can be found at Section 3.1.1, Section 3.1.2 and Section 3.1.4.



(a) Results of experiment 1 and 2



(b) Results of experiment 3



(c) Results of experiment 4

Figure 5.7: Mean average precision with different sub-images sizes for the Wang dataset (AP@20).

Evaluation measures used for the evaluation of the system are precision, recall, precision at n and ANMRR using equation 3.2, 3.1, 3.3, 3.4 and 3.9 respectively. R-precision was used evaluate the system on Flickr25K dataset.

AP@n was calculated for the Wang and Oliva and Torralba datasets. AP@20, AP@50 and AP@100 were calculated for each class to compare with existing systems.

5.4.2 Selection of Appropriate Block Sizes

As the primary contribution in this work was to provide precise image decomposition method, the precision of different decomposition methods were tested. Four different experiments were conducted as follows:

Experiment 1: Different sub-image sizes were tested. Sub-image size 16x16, 32x32, 64x64, and 128x128 were tested and found that 16x16 and 32x32 (MAP = 0.575 and 0.602 respectively) resulted in inferior performances compared with other two. The reason may be that an image size less than 64x64 may be not enough to extract features. The sub-image size 64*64 provides the best overall. Sub-image size 128x128 (MAP=0.599) resulted in an inferior performance when compared with 64x64 (MAP = 0.654). This may be due to lack of information to differentiate images as available data is not enough to generate descriptive vocabularies. These experiments concluded that image decomposition with the size of a 64*64 sub-image is the best to use in a grid-based CBIR among selected sizes. Results of this experiment on the Wang dataset are demonstrated in the first four columns of the figure 5.7.a

Experiment 2: As it was found that 64x64 is the best among all the tested sub-image sizes, overlapping sub-image generation was tested by moving a window of 64x64. It showed promising performance in retrieval quality with the MAP of 0.71, which substantially transcended that of the non-overlapping 64*64 sized sub-images (MAP of 0.654). Result of this experiment on the Wang dataset is demonstrated in the last column of the figure 5.7.a.

Experiment 3: Without stopping at this stage, grid-based image decomposition was tested. In this experiment images were sub-divided in a 2 by 2 grid, 3 by 3 grid, 4 by 4 grid and 5 by 5 grid. 2 by 2 and 4 by 4 are same as 128x128 and 64x64. Among these image partitioning methods, 3 by 3 grid-based image

sub-division provided the best performance in the retrieval quality by gaining a MAP of 0.727 and it is computationally effective than overlapping of sub-images. Results of this experiment on the Wang dataset are demonstrated in figure 5.7.b.

Experiment 4: The proposed circular image decomposition method which was described in Section 5.2.3 was tested. Although it performed well (MAP = 0.684), it did not perform as expected. But the MAP of this method only inferior to the methods of image decomposing by 3 by 3 grid and 64x64 overlapping. We assumed that this works more effectively with objects as it gave a superior performance in some classes, like faces and dinosaurs. Result of this experiment on the Wang dataset is demonstrated in figure 5.7.c.

From all of the image decomposition methods, 3 by 3 grid-based image decomposition gave the best results (MAP = 0.727). From the results of the experiments it was concluded that image sub-division to 3 by 3 blocks is the best way to partition. From this we could say it is better to use a sub-image size around 80x80 for feature extraction. It must be mentioned that all the images had been taken to 256x256 size during these experiments. MAP of these image portioning methods on the Wang dataset is as shown in figure 5.7. All the image signatures are of size 1024 bits and these experiments were done before the feature selection (unoptimised feature set), as the final target was to find the most suitable sub-image size.

5.4.3 Selection the Most Suitable Vocabulary Size

Since this CBIR system adapted BoW representation, visual vocabularies should be made from the extracted features. Therefore, independent visual vocabularies were generated for each feature. Although a large vocabulary tends to improve retrieval accuracy, it does not mean that a larger vocabulary definitely leads to a higher retrieval accuracy. We noted that the retrieval accuracy first increases, then reaches its best and drops with the increasing vocabulary size. Moreover, vocabularies were fairly small, as the representation was dense. Different vocabulary sizes were used for the experiment in Section 5.4.2 as we used the most suitable vocabulary for each case. We only describe the vocabulary generation for the selected representation. After experimenting with different vocabulary sizes, 20 was selected as the vocabulary size for the Wang and Oliva and Torralba datasets.

Therefore , the full vocabulary size was 140, as we used seven features. Figure 5.8 shows the average precision variation with the change of vocabulary size.

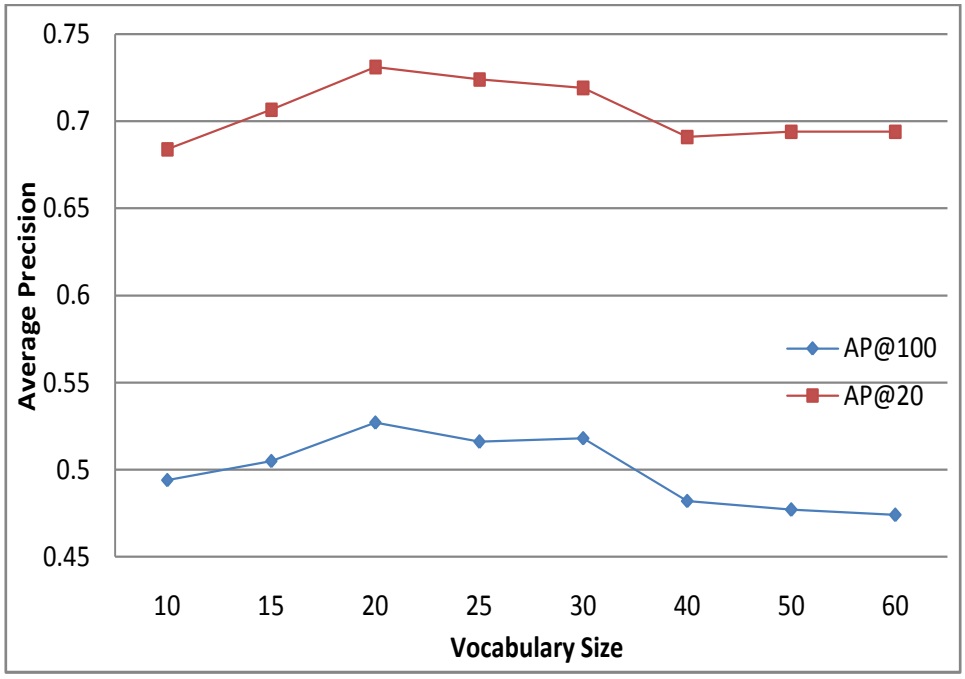


Figure 5.8: Mean average precision for different vocabulary sizes on the Wang dataset.

When the size of the dataset increases, the vocabulary size increases in log scale, and initially without further experimentation with vocabulary sizes for each dataset we forecasted the vocabulary sizes and . Therefore, we used 40 for the Corel, Caltech and Flickr databases and 80 for the SUN database. Finally it was found that the selected values are suitable for those datasets by further experiments.

5.4.4 The Effect of Term Statistics on Effectiveness

In text retrieval, each document is represented by a vector of word frequencies and apply weights on the vector which depend on the term statistics. As this research work used BoW representation, the effect of term statistics on the performance was evaluated. We used well known term frequency - inverse document frequency (TF-IDF) [Salton et al., 1975], Okapi BM25 (BM25) [Manning et al., 2008] and Log-likelihood (LL) [Chappell, 2015] by using equations 3.10, 3.11 and 3.12

Table 5.1: Average Precision (AP) at n with different term statistics on the Wang dataset with a 1024 bits signature size (AP@n).

AP@n	No Stat	BM25	LL	TF-IDF
20	0.803	0.795	**0.806**	0.796
100	0.589	0.577	**0.597**	0.584

Table 5.2: Average Precision (AP) at n with different term statistics on the Wang dataset with an 8192 bits signature size (AP@n).

AP@n	No Stat	BM25	LL	TF-IDF
20	0.826	0.813	**0.827**	0.812
100	0.618	0.610	**0.625**	0.606

Table 5.3: Average Precision (AP) at n with different term statistics on the Oliva and Torralba dataset with a 1024 bits signature size (AP@n).

AP@n	No Stat	BM25	LL	TF-IDF
20	0.758	0.741	**0.766**	0.762
50	0.709	0.685	**0.713**	0.711
100	0.636	0.616	**0.645**	0.644

Table 5.4: Average Precision (AP) at n with different term statistics on the Oliva and Torralba dataset with an 8192 bits signature size (AP@n).

AP@n	No Stat	BM25	LL	TF-IDF
20	0.775	0.751	**0.776**	0.771
50	0.725	0.704	**0.731**	0.730
100	0.650	0.631	0.658	**0.659**

respectively for the experiments.

Experiments were carried out on Wang and Oliva and Torralba datasets. Table 5.1 and table 5.2 demonstrate performance by applying different term statistics to the Wang dataset. Furthermore, table 5.3 and table 5.4 demonstrate perfor-

mance by applying different term statistics to the Oliva and Torralba dataset. According to the tables, there is an improvement with TF-IDF and Log-likelihood. Furthermore, Log-likelihood outperforms TF-IDF and Okapi BM25. Therefore, Log-likelihood works better for these general image datasets. However, according the results, the increase in performance is only small. This may be because the vocabulary is fairly small and there are not many frequent words in the documents. As the improvement is fairly small we did not use term statistics in the system as we have to sacrifice efficiency to gain this effectiveness. Nevertheless, this is useful when the database size is getting bigger.

5.4.5 Selection of the Most Suitable Image Signature Size

Signature size provides a trade-off between retrieval quality of retrieval efficiency. Thus, signature size can be chosen based on the retrieval speed and quality. A higher or smaller signature size can be selected based on the targeted parameter of retrieval quality or speed respectively. Hence, a medium-sized signature ensured the trade-off between the retrieval quality and speed. Different signature sizes were tested on the Wang dataset in the early stages of this research to study the behaviour of different feature settings which were not optimised [Chathurani et al., 2014]. The same signature sizes were tested on the new feature selection, which is described in Section 5.2.1, and it is shown in figure 5.9. To confirm it, the same test was run on the Oliva and Torralba dataset with an un-optimised (before feature selection) feature setting, as well as an optimised (after feature selection) feature setting.

Table 5.5 and table 5.6 provide a detail explanation. Those tables show how the AP and MAP varies with the signature sizes on the Wang dataset. Even a 64 bits signature size achieves more than 50% (MAP = 0.55) precision, which is considerable. Moreover, signatures of 4096 bits - 8192 bits in size achieve the highest effectiveness. 4480 bits were necessary to represent a typical histogram with these features and the representation and the proposed approach had achieved similar performance in quality to the histogram-based results even with a 1024 bits signature size (histogram-based -> MAP = 0.807 and signature-based with the size of 1024 bits -> MAP = 0.803).

From these experiments it could be concluded that variation in the performance

of the retrieval quality with the signature size has a same pattern. Precision was increased to a certain signature size and then started to decrease the performance. From this it could be concluded that a large signature size will not provide a higher performance always, because long signatures may have overlaps at a particular length. They also require additional processing time for signature generation and searching. In this thesis initially 1024 bits signature size was used for experiments. However, a 8192 bits signature size was used in CBIR-ISIG system as it provides good compromise between retrieval quality and speed (retrieval speed will be discussed in Chapter 8). Average precision against signature size in figure 5.10 shows how precision is increased with the signature size for several datasets and there cannot be seen any significant improvement after 4096-8192 bits signature sizes which suggesting that our choice is the most appropriate.

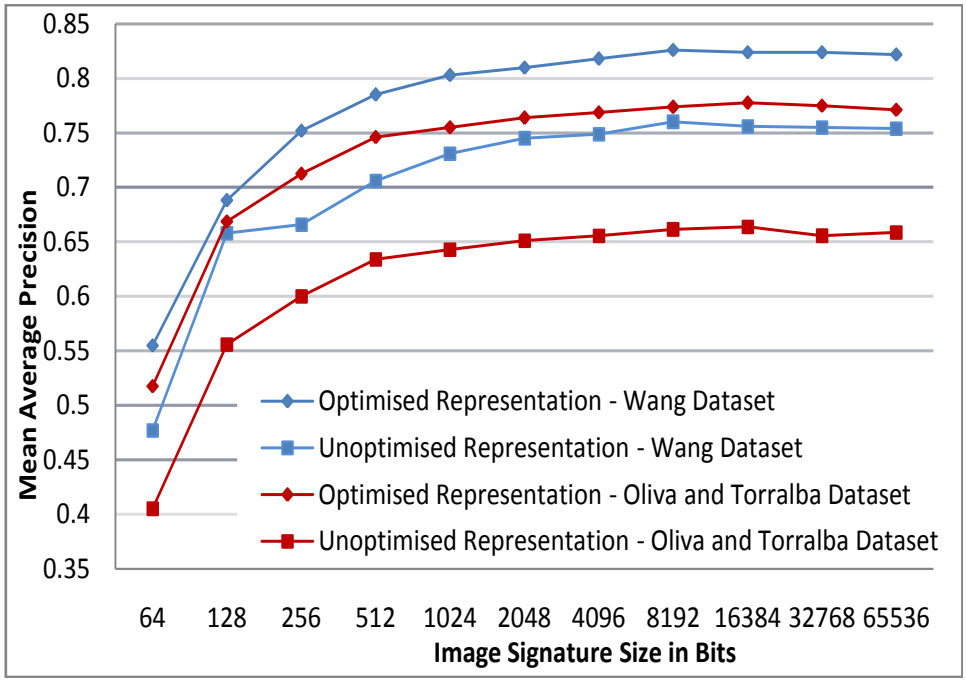


Figure 5.9: Retrieval effectiveness vs. signature size over two datasets (AP@20). Signatures of 4K-8192 bits in size achieve the highest effectiveness.

Table 5.5: Average Precision (AP) of each class along with whole dataset with different signature size (AP for the top 20 images) for the Wang dataset-before feature selection.

Class	Signature Size (in bits)										
	64	128	256	512	1024	2048	4096	8192	16384	32768	65536
Africans	0.40	0.55	0.64	0.70	0.70	0.73	0.73	0.75	0.75	0.74	0.74
Beach	0.31	0.47	0.51	0.59	0.61	0.63	0.65	0.64	0.64	0.64	0.64
Building	0.35	0.38	0.40	0.40	0.46	0.48	0.47	0.50	0.48	0.48	0.48
Bus	0.55	0.76	0.77	0.80	0.81	0.84	0.84	0.85	0.84	0.84	0.84
Dinosaur	0.81	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Elephant	0.31	0.59	0.57	0.62	0.65	0.69	0.69	0.70	0.70	0.70	0.70
Flower	0.71	0.90	0.95	0.94	0.94	0.94	0.94	0.95	0.95	0.95	0.95
Horse	0.65	0.83	0.88	0.91	0.91	0.92	0.93	0.94	0.93	0.93	0.93
Mountain	0.29	0.46	0.40	0.46	0.58	0.53	0.56	0.58	0.58	0.58	0.58
Food	0.32	0.55	0.54	0.64	0.65	0.69	0.68	0.69	0.69	0.69	0.68
Average Precision	0.477	0.658	0.666	0.706	0.731	0.745	0.749	0.76	0.756	0.755	0.754

Table 5.6: Average Precision (AP) of each class along with whole dataset with different signature size (AP for the top 20 images) for the Wang dataset -after feature selection.

Class	Signature Size (in bits)										
	64	128	256	512	1024	2048	4096	8192	16384	32768	65536
Africans	0.48	0.60	0.64	0.68	0.71	0.71	0.73	0.75	0.75	0.74	0.73
Beach	0.50	0.54	0.58	0.61	0.65	0.65	0.66	0.67	0.68	0.68	0.67
Building	0.41	0.47	0.52	0.49	0.51	0.51	0.53	0.53	0.53	0.52	0.53
Bus	0.40	0.58	0.75	0.82	0.85	0.85	0.85	0.86	0.85	0.85	0.85
Dinosaur	0.80	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Elephant	0.55	0.76	0.78	0.86	0.81	0.88	0.86	0.86	0.85	0.87	0.86
Flower	0.74	0.92	0.91	0.95	0.96	0.96	0.98	0.98	0.98	0.97	0.98
Horse	0.64	0.86	0.92	0.94	0.95	0.97	0.97	0.98	0.97	0.97	0.97
Mountain	0.30	0.45	0.63	0.67	0.74	0.73	0.75	0.77	0.78	0.78	0.77
Food	0.69	0.74	0.81	0.83	0.85	0.84	0.85	0.86	0.85	0.86	0.86
Average Precision	0.551	0.69	0.754	0.785	0.803	0.81	0.818	0.826	0.824	0.824	0.822

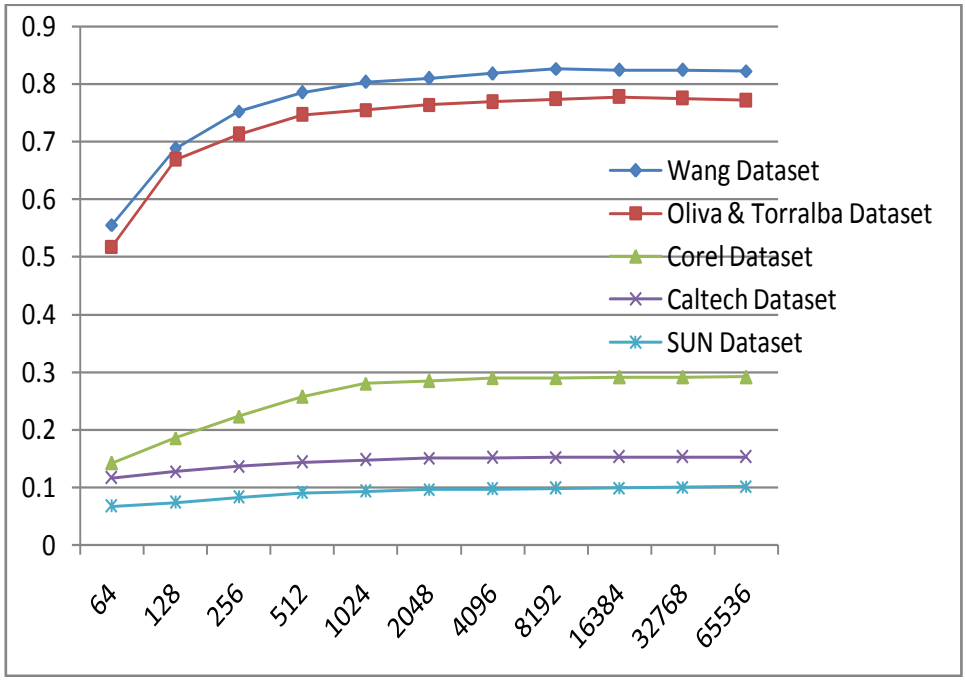


Figure 5.10: Retrieval effectiveness vs. signature size (AP@20).

5.4.6 The Ability of Signatures in Preserving Similarity

In order to investigate this, some experiments were performed. A seed is used to generate pseudo-random signatures during random projection. Different seed values will result in different image signatures for the same image. Therefore, the hamming distance will be different when searching signatures that were generated with different seed values. Nevertheless, random indexing is a form of topology preserving hashing. Hence, images that are similar will have similar signatures even under different random seeds, and conversely for dissimilar images. Figure 5.11, figure 5.12 and figure 5.13 depict the mutual distance matrix of signatures as heat maps that show the distance relationship between images. These heat maps show the hamming distance from each image to all other images. The image signatures are grouped by class and the diagonal shows self similarity. All images of the same class are adjacent. Here, 100 images from the Wang dataset were used to compute the distance matrix and it comprised of 10 images from each class. The lighter colour (light yellow) indicate larger hamming distances while the brighter colour (red) indicates lower hamming distances corresponding to similar images.

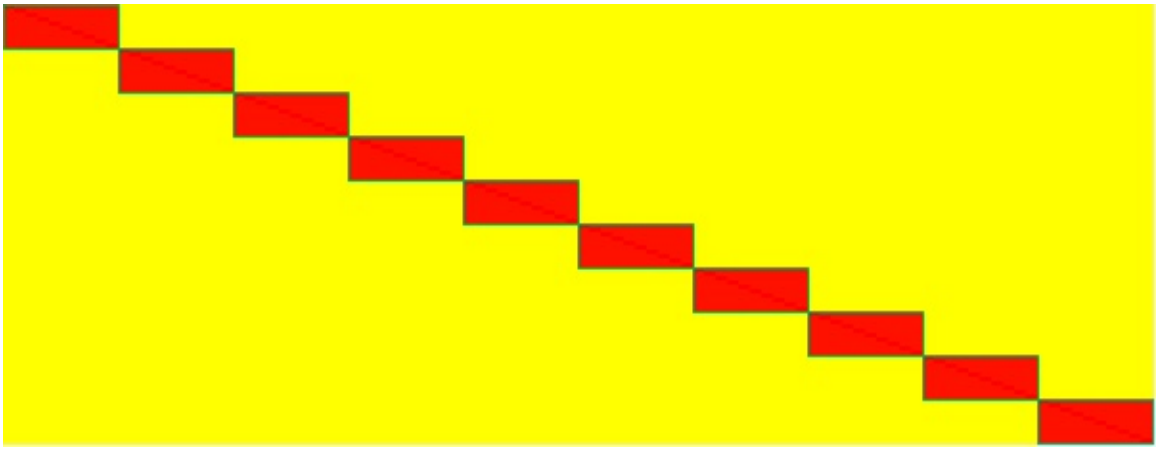
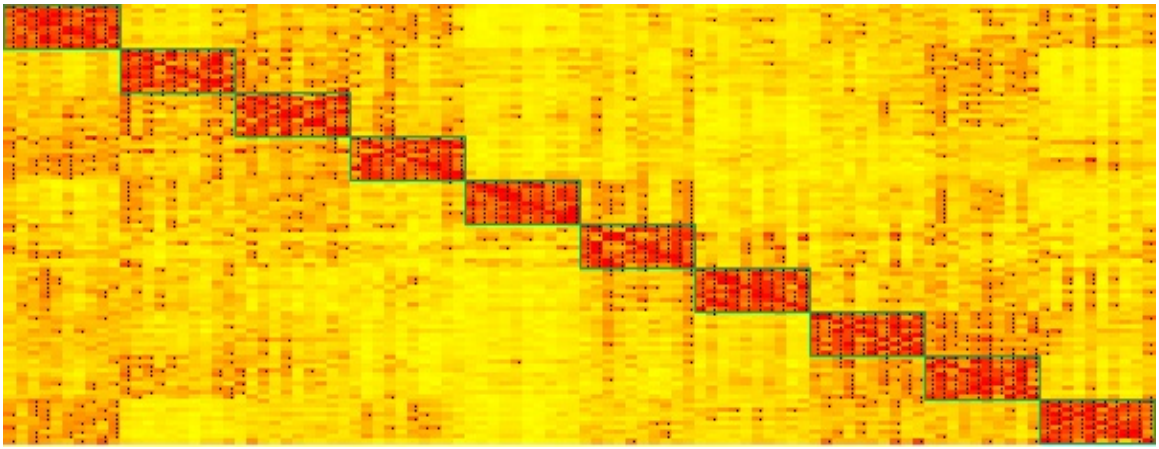
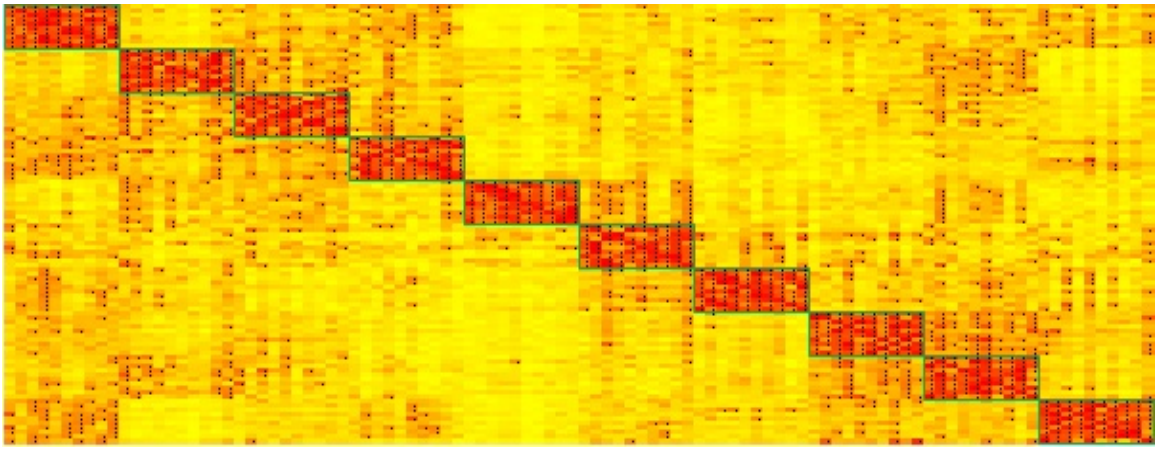


Figure 5.11: Ideal case - Heat map showing the hamming distance for all the similarities for 100 images of (10 image from each class).

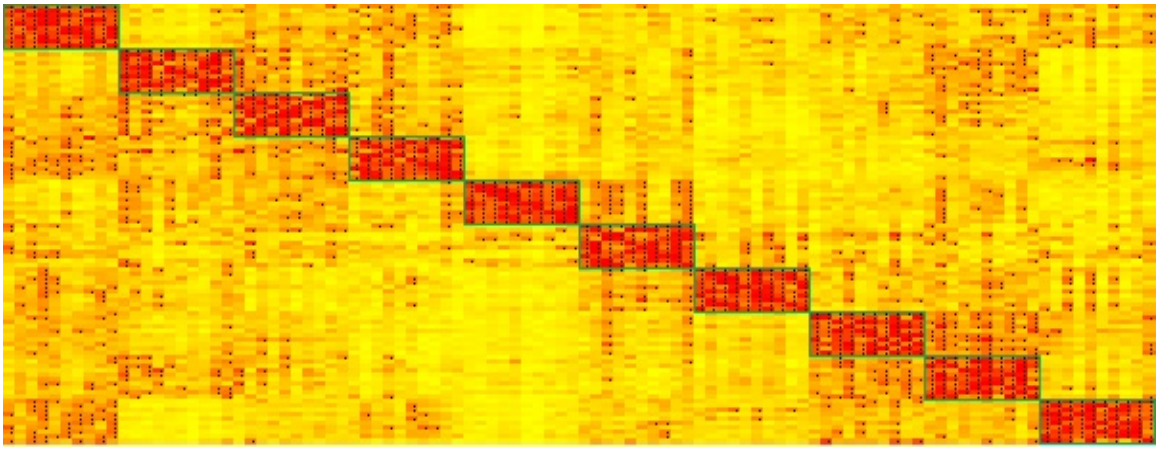
Figure 5.11 demonstrates an ideal case where signatures within each class are identical and signatures from different classes are uncorrelated, while figure 5.12 demonstrates real values from experiments, using three different random number generator seeds (This image set comprised of images which have high inter-class variability and less intra-class variability) and figure 5.13 shows a heat map for another 100 images which have low inter-class variability and high intra-class variability. In this figure, each class is highlighted by bounding box for assisting recognition. Generally intra-class distance is smaller than inter-class distance which means that images that are in the same class have signatures that are closer, than the images outside the class. This can be seen from figure 5.12 and figure 5.13. There is a block pattern along the diagonal as expected from the ideal case in figure 5.11. Different seeds will generate different signatures. But regardless of the seed the intra-class distance and inter-class distance have remained largely separated (figure 5.12.a, 5.12.b, 5.12.c have similar pattern). Classes are quite well separated as demonstrated in the figures although there is still residual confusion. This is due to the feature overlap. As images share image features, they are not disjoint resulting in overlap of image signature as can be seen in figure 5.12 and figure 5.13. While we have not optimised feature selection or engineering, since it was deemed to be out of scope for this research, we have applied a standard wrapper-based feature selection as described in Section 5.2.1.



(a) Seed 0



(b) Seed 1



(c) Seed 2

Figure 5.12: Heat maps for averagely good results showing the hamming distance for three different seeds. Hamming distance for all the similarities for 100 images of Wang dataset (10 image from each class).

Figure 5.12 and figure 5.13 demonstrate that the signatures obtained preserve similarity well. The figures demonstrate that complete process, starting from feature extraction, through BoW generation and random projection, and, finally binary image signatures, leads to a representation that is discriminative enough to support the task of image retrieval by preserving similarly between images.

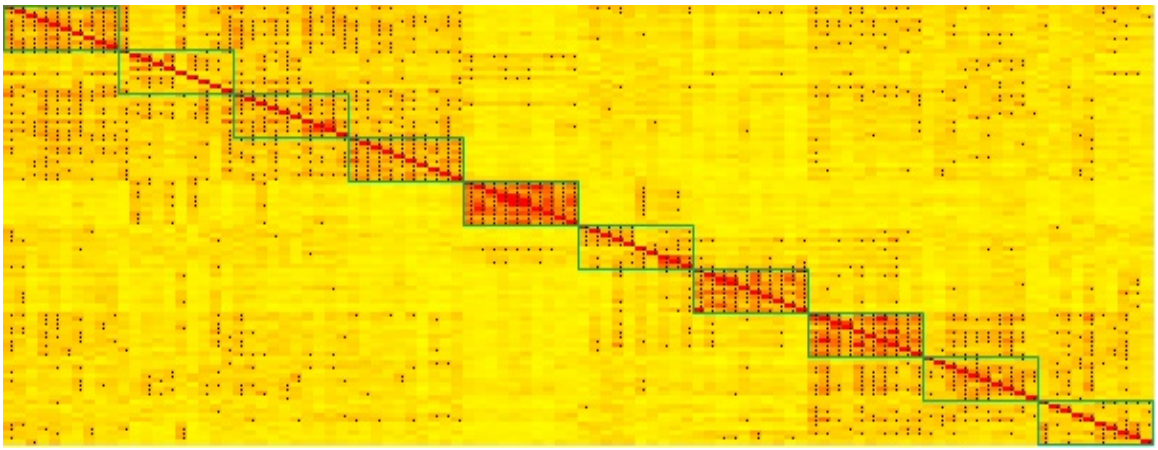


Figure 5.13: Heat map for set of queries which have low inter-class variability and high intra-class variability (100 images of Wang dataset, 10 image from each class).

5.4.7 Retrieval Performance

Retrieval quality of the proposed image retrieval approach was tested on all datasets. MAP@N (N is the number of images retrieved at a time) was calculated for the Wang and Oliva and Torralba dataset starting from N=5 until N=100 and it is shown in figure 5.14. MAP@100 is 0.61 and 0.64 for the Wang and Oliva and Torralba dataset respectively with the 1024 bits signature size. Both the datasets achieved more than 50% accuracy in retrieval which is great as a measure.

As the Oliva and Torralba dataset has different class sizes, the retrieval quality depends on the class sizes. Therefore, an ANMRR measure was calculated for both the datasets and it is shown in table 5.7. If the ANMRR value is near to zero, then the system has good potential to retrieve correct images irrespective of the size of ground truth. ANMRR = 0.1638 and 0. 2333 means that the proposed approach has good potential.

R-Precision was calculated for the Flickr25K dataset and average R-precision

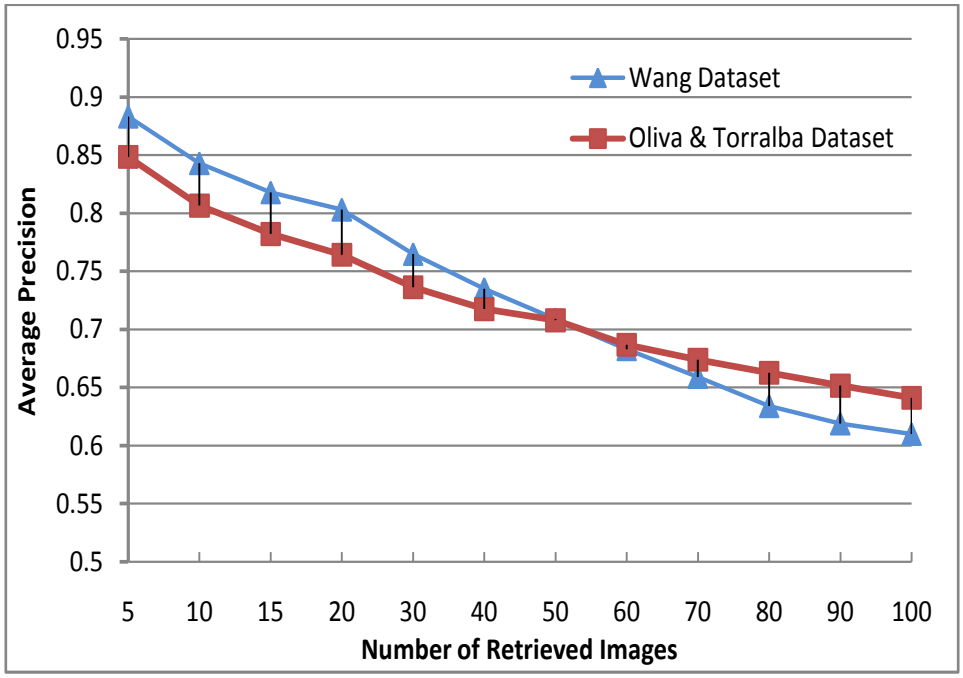


Figure 5.14: Mean average precision at N for the Wang and Oliva and Torralba datasets with 1024 bits signature size.

Table 5.7: ANMRR measure of the proposed approach for the Wang and Oliva and Torralba datasets.

Dataset	ANMRR
Wang Dataset	0.1638
Oliva & Torralba Dataset	0.2333

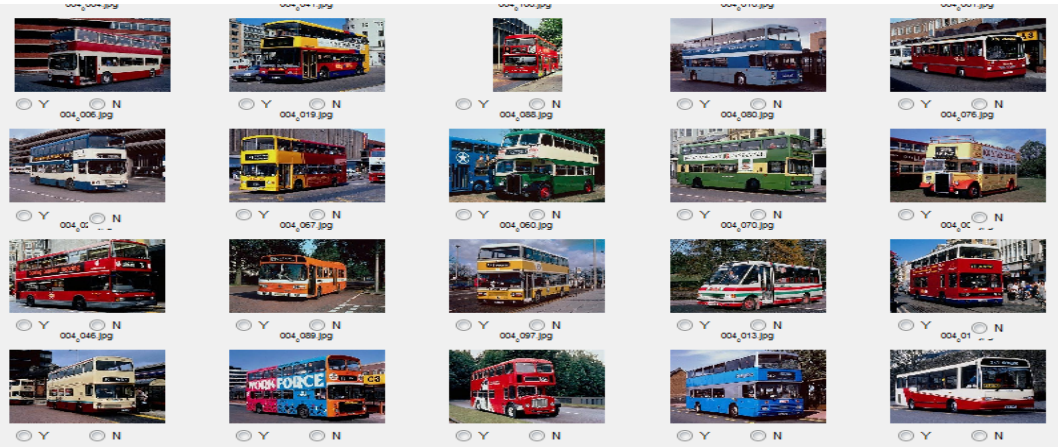
was calculated at the end and it is 0.284 and R-precision for each class, as shown in table 5.8. It shows the coverage of each class as a percentage and the number of images in each class along with the R-precision. Here R is the number of relevant images to the topic. We used R-precision for this dataset as the system could easily achieve high values for MAP@20 (100%) and MAP@100 (99.49%), as the class sizes are bigger and the classes have many overlaps. Thus, there is a greater possibility to get correct images in first of the top-ranked list itself.

Table 5.8: R-precision of the Flickr 25K dataset. Here most of the images are classified in several classes.

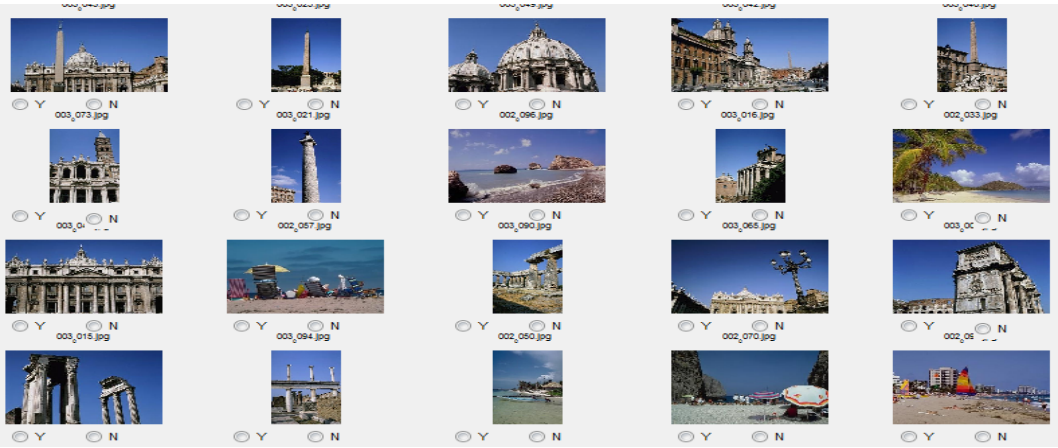
Category	Coverage from whole set (25000)	No Of Images	R-Precision
animals	13%	3216	0.26
baby	1%	259	0.06
baby_r1	0%	116	0.05
bird	3%	742	0.1
bird_r1	2%	484	0.07
car	5%	1177	0.19
car_r1	2%	380	0.08
clouds	15%	3700	0.37
clouds_r1	5%	1350	0.21
dog	3%	684	0.1
dog_r1	2%	590	0.09
female	25%	6184	0.48
female_r1	16%	3982	0.38
flower	7%	1823	0.17
flower_r1	4%	1077	0.1
food	4%	990	0.14
indoor	33%	8313	0.45
lake	3%	791	0.15
male	24%	6081	0.48
male_r1	15%	3647	0.38
night	11%	2711	0.5
night_r1	3%	669	0.24
people	41%	10373	0.59
people_r1	31%	7849	0.53
plant_life	35%	8763	0.59
portrait	16%	3931	0.37
portrait_r1	15%	3829	0.36
river	4%	894	0.18
river_r1	1%	149	0.05
sea	5%	1322	0.22
sea_r1	1%	214	0.08
sky	32%	7912	0.52
structures	40%	9992	0.68
sunset	9%	2135	0.28
transport	12%	2895	0.35
tree	19%	4683	0.43
tree_r1	3%	668	0.12
water	13%	3331	0.38
Average Precision			0.284

The top 20 retrieved results for some queries are shown in the figure 5.15, covering Wang dataset and Oliva and Torralba dataset. In those images, almost all are correct matches, in the first 20 except for one or two mismatches. These results are for 1024 bits signature size. From this, we can see that even with the 1024 bits signature size it has achieved a good performance. Up to now, we used the 1024 bits signature size for evaluation and we selected a 8192 bits signature size as we required to maintain a good retrieval performance while maintaining the speed for the CBIR-ISIG system.

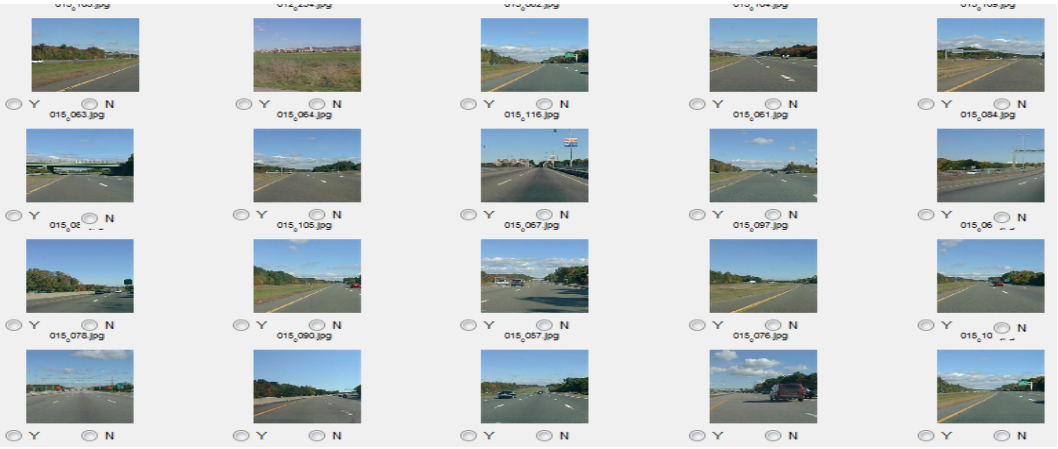
Even though evaluations were carried out on different datasets to highlight the system's effectiveness as above, we compared our systems only against other state of the art solutions for which evaluation results were available on the same datasests used here for evaluation, i.e. Wang, Corel, Oliva nd Torralba. AP was calculated at the first 20 and the first 100 for the Wang dataset and at 50 for the Oliva and Torralba dataset in the comparison. We used these measures specifically to compare against available systems as those systems have used AP@20, AP@50 and AP@100 to evaluate retrieval performance. Retrieval quality was compared with existing systems, namely block-based retrieval using local binary patterns [Takala et al., 2005], block-based retrieval using a combination of colour, texture and shape features [Hiremath and Pujari, 2007a], CBIR based on combination of colour, texture and spatial features [Lin et al., 2009] SIFT-LBP [Yuan et al., 2011b], retrieval based on colour and shape features [Saad et al., 2011], histogram-based image retrieval [Mansoori et al., 2013], retrieval using colour, texture and shape features with the support of SVMs [Hiwale et al., 2015] and CBIR with the combination of both global and local descriptors [Douik et al., 2016] for AP at 20 on Wang. Moreover, AP at 100 on Wang was compared with SIMPLICITY [Li et al., 2000], FIRM [Chen and Wang, 2002], image retrieval using salient points (salient points detected by Harris Comer Detector (SP by HCD), colour salient points (CSP)) [Hiremath and Pujari, 2008], edge based retrieval [Banerjee et al., 2009] and retrieval using ripplet transform [Chowdhury et al., 2012]. Additionally retrieval quality on the Oliva and Torralba dataset was compared with a system which has used the bag of regions for retrieval [Gokalp and Aksoy, 2007] for AP at 50. These systems were selected due to the fact that those systems have used Wang and Oliva datasets for evaluation.



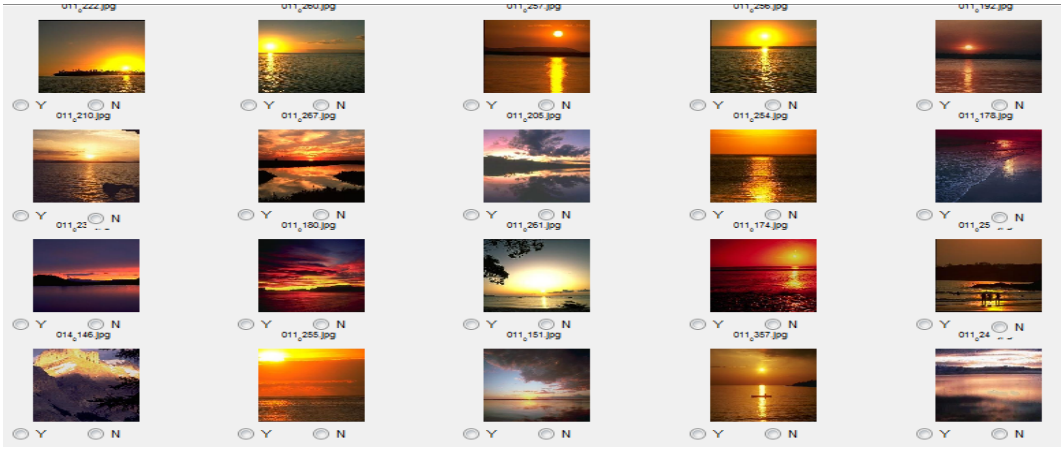
(a) Buses



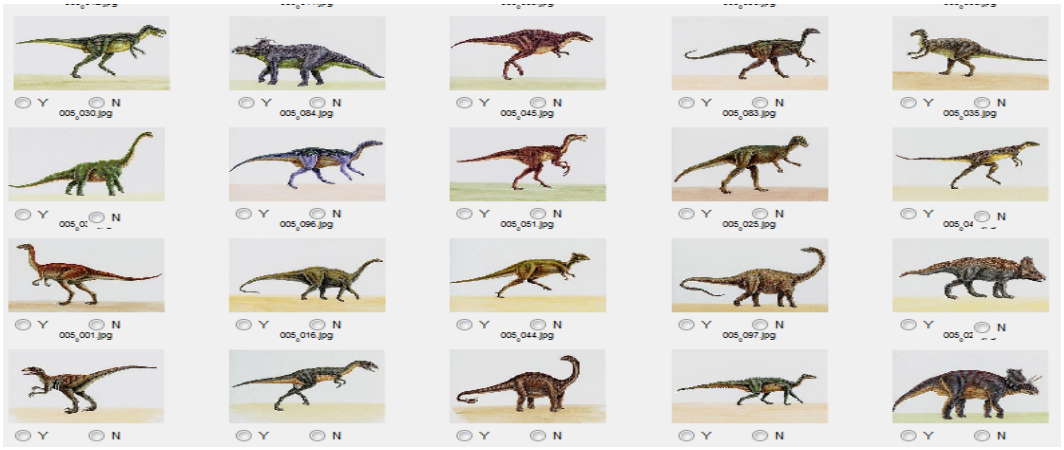
(b) Buildings



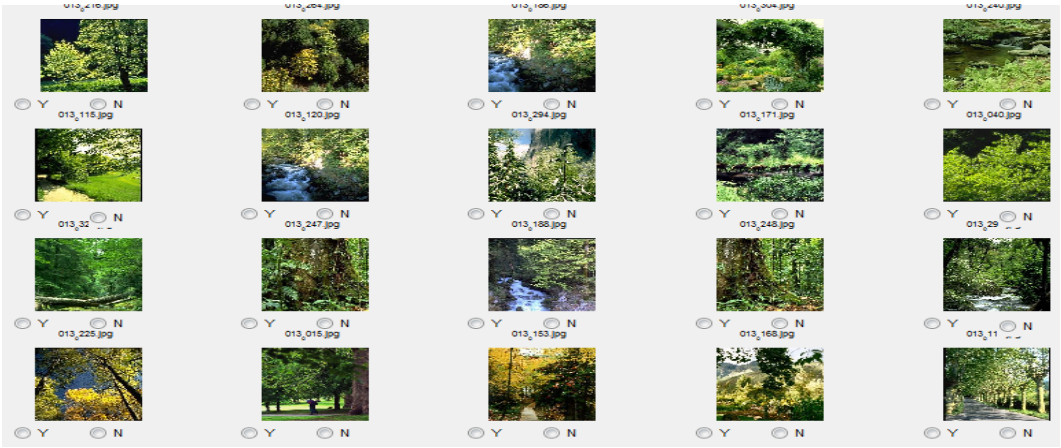
(c) Highway



(a) Beach



(b) Dinosaurs



(c) Forest

Figure 5.15: The top 20 images covering some queries in the Wang dataset and Oliva and Torralba dataset with the 1024 bits signature size (top left is the query image.) [Chathurani et al., 2015a]

Table 5.9 and table 5.10 shows the performance comparison with the existing systems for AP@20 and AP@100 based on the Wang dataset. Table 5.11 shows the performance comparison with an existing region-based approach for MAP@50 on the Oliva and Torralba dataset. The highest value of each class from the compared systems are bolded in all the tables.

When we consider the results in table 5.9, table 5.10 and table 5.11 significant test cannot be computed to see the performance improvement over the other systems as results for each query is not given instead mean average precision of each class. In table 5.9 the proposed CBIR-ISIG system (last column) shows the highest average precision for half of the classes and generally averages are higher. According to the table 5.10 the proposed CBIR-ISIG system (last column) shows the highest average precision for most of the classes among compared results. Table 5.11 demonstrates that CBIR-ISIG system (last column) have the highest average precision for most of the classes.

Compared results show that the proposed approach has a high ability to retrieve correct images, as it shows much improved MAP over the other systems. MAP of the proposed system is the highest for all cases (MAP@20 = 0.83 and MAP@100 = 0.62 for the Wang dataset and MAP@50 = 0.73 for the Oliva and Torralba dataset) [Chathurani et al., 2015a, Chathurani et al., 2016a].

Table 5.9: Average Precision (AP) of each class along with whole dataset for the Wang dataset compared with the performance of the systems in the literature (AP@20). Here ‡ means that results are statistically significantly greater than 99% significance level when comparing against un-optimised feature setting (column before the last column).

Class	2005 [Takala et al., 2005]	2007 [Hiremath and Pujari, 2007a]	2009 [Lin et al., 2009]	2011 [Yuan et al., 2011b]	2011 [Saad et al., 2011]	2013 [Mansoori et al., 2013]	2014 [Hiwale et al., 2015]	2016 [Douik et al., 2016]	CBIR-ISIG [Chathurani et al., 2014]	CBIR-ISIG [Chathurani et al., 2015a]
Africans	0.23	0.48	0.68	0.57	**0.90**	0.68	0.76	0.70	0.75	0.75‡
Beach	0.23	0.34	0.54	0.58	0.38	0.28	0.54	0.45	0.64	**0.68**‡
Building	0.23	0.36	0.56	0.43	**0.72**	0.56	0.67	0.59	0.50	0.53‡
Bus	0.23	0.61	0.88	0.93	0.49	0.84	0.84	**0.98**	0.85	0.86
Dinosaur	0.23	0.95	0.99	0.98	**1.00**	0.81	**1.00**	**1.00**	**1.00**	**1.00**
Elephant	0.23	0.48	0.66	0.58	0.39	0.58	0.70	0.81	0.64	**0.86**‡
Flower	0.23	0.61	0.89	0.83	0.56	0.55	0.95	**1.00**	0.96	0.98
Horse	0.23	0.74	0.80	0.68	0.87	0.87	0.94	0.90	0.93	**0.98**‡
Mountain	0.23	0.42	0.52	0.46	0.45	0.48	0.58	0.52	0.59	**0.77**‡
Food	0.23	0.50	0.73	0.53	**0.87**	0.66	0.68	0.61	0.69	0.86‡
Average Precision	0.23	0.549	0.725	0.657	0.663	0.633	0.764	0.751	0.76	**0.827**

Table 5.10: Average Precision (AP) of each class along with whole dataset for the Wang dataset compared with the performance of the systems in the literature (AP@100). Here ‡ means that results are statistically significantly greater than 99% significance level when comparing against un-optimised feature setting (column before the last column).

Class	2000 [Li et al., 2000]	2002 [Chen and Wang, 2002]	2008 [Hiremath and Pujari, 2008]	2008 [Hiremath and Pujari, 2008]	2009 [Banerjee et al., 2009]	2012 [Chowdhury et al., 2012]	CBIR -ISIG [Chathurani et al., 2014]	CBIR -ISIG [Chathurani et al., 2015a]
Africans	0.48	0.47	0.40	0.48	0.45	0.49	0.50	**0.52**‡
Beach	0.33	0.33	0.31	0.34	0.35	0.40	0.45	**0.46**‡
Building	0.33	0.33	0.32	0.33	0.35	**0.39**	0.33	0.38‡
Bus	0.36	0.60	0.44	0.52	0.60	0.58	0.62	**0.63**‡
Dinosaur	0.98	0.95	0.92	0.95	0.95	0.96	0.98	**0.98**‡
Elephant	0.40	0.25	0.28	0.40	**0.60**	0.50	0.44	0.57‡
Flower	0.40	0.63	0.58	0.60	0.65	0.75	0.75	**0.84**‡
Horse	0.72	0.63	0.68	0.70	0.70	**0.80**	0.68	0.76‡
Mountain	0.34	0.25	0.32	0.36	0.40	0.40	0.36	**0.47**‡
Food	0.34	0.49	0.44	0.46	0.40	0.51	0.41	**0.60**‡
Average Precision	0.468	0.493	0.469	0.514	0.545	0.578	0.552	**0.621**

Table 5.11: Average Precision (AP) of each class along with whole dataset (Oliva and Torralba) with performance in the literature (AP@50).

Class	Region-based [Gokalp and Aksoy, 2007]	CBIR-ISIG [Chathurani et al., 2016a]
Coast (beach)	**0.84**	0.63
Country side	**0.50**	**0.50**
Forest	0.76	**0.85**
Mountain	**0.80**	0.73
Highway	0.62	**0.75**
Street	0.44	**0.94**
City centre	0.38	**0.67**
Tall buildings		0.73
Average Precision	0.62	**0.73**

5.5 Application

5.5.1 Content-Based Image (Object) Retrieval with Rotational Invariant Bag-of-Visual Words Representation

This research work tried to achieve rotational invariant feature representation using global descriptors. This section describes the proposed Rotation Invariant Bag of Words (RIBoW) model using circular image decomposition. This divides each object into similar sized parts starting from the centre by assuming that all the objects are in the centre of an image. Objects stay in the middle of an image in most cases. In some cases that this assumption will not work and this can be confirmed by looking at figure 5.22, figure 5.23 and figure 5.20.

Unlike spatial pyramid matching (SPM), this method used signature-based representation which can be extended this work for large scale datasets. This proposed approach was evaluated using two standard datasets. This will work especially when objects stay in the middle and is suitable for nearest neighbour

detection in those kind of images as it has good early precision.

5.5.2 Background Work

During the past decade, the BoW approach has achieved popularity in the fields of classification (object, scene) and retrieval (image, video) in CBIR [Sivic and Zisserman, 2003, Lazebnik et al., 2006, Aman et al., 2010, Rahat et al., 2012, Yuan et al., 2011b, Csurka et al., 2004, Torralba et al., 2008a] because of its simplicity and good relative performance. This approach was introduced by Sivic and Zisserman [Sivic and Zisserman, 2003] to the computer vision community and was inspired by BoW model in text document retrieval. The BoF approach is analogous to BoW representation in text document retrieval. BoW representation is also suitable for large databases as it scales efficiently to large collections, and the approach is flexible with geometry deformations and viewpoints and it provides vector representation for sets. Finally, it provides a compact summary of image content. In the BoW approach in CBIR, the visual vocabulary or visual codebook is formed by clustering image features that are extracted from images in the database. Firstly, similar features are gathered together where each cluster centre stands for a visual word. After that, feature vectors are mapped to those visual words and each image is represented as a histogram of visual words which provides the occurrence of each word that appears in an image. Though BoW has these advantages, it has its disadvantages too, as it discards spatial information which severely affects the retrieval performance.

Spatial location is useful in region classification. Even though spatial location is simply defined as top, bottom, left and right, it is still important to differentiate, for an example, sky from sea. But their colour and texture may be same. Spatial information can be included in BoF by spatial pyramid structure to improve the efficiency of the BoF approach and it has gained superior performance in classification and retrieval applications [Lazebnik et al., 2006, Yang et al., 2009]. A spatial pyramid is a collection of order-less feature histograms computed over cells defined by multi-level recursive image decomposition. At level 0, the decomposition consists of just a single cell, and the representation is equivalent to a standard bag of features. At level 1, the image is subdivided into four quadrants, yielding four feature histograms, and so on [Yang et al., 2009]. It is found that for strong features,

image subdivision till level 2 is enough and there is no performance improvement in level 3 because it is too finely subdivided and it yields too few matches [Lazebnik et al., 2006]. Figure 5.16 shows the spatial pyramid representation for two levels.

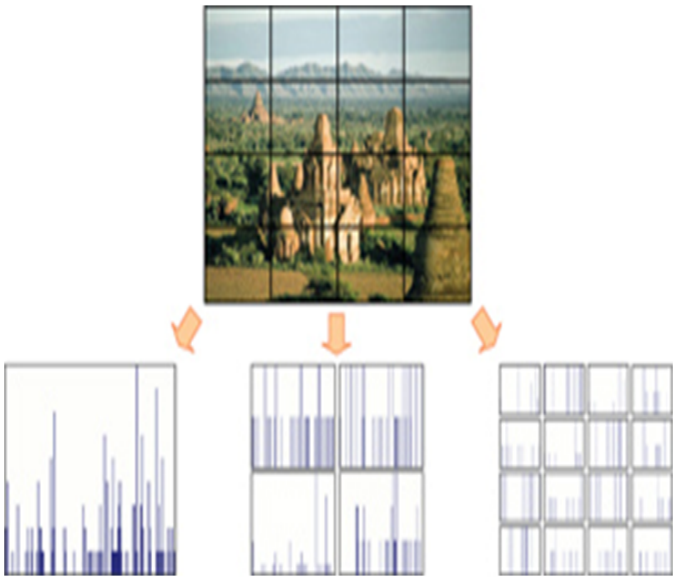


Figure 5.16: A schematic illustration of the spatial pyramid representation.

Finally, feature vector is generated by combining all the histograms in different levels and it is named as the Pyramid Histogram Of visual Words (PHOW). Pyramid matches works by placing a sequence of increasingly coarser grids over the feature space and taking a weighted sum of the number of matches that occur at each level of resolution. At any fixed resolution, two points are matched if they fall into the same cell of the grid. Matches found at finer resolutions are weighted more highly than matches found at coarser resolutions because higher levels provide more precise representation than lower levels [Yang et al., 2009]. However, the SPM cannot handle the translation, rotation and scale variance of an image, though it encodes spatial information.

Recently, a rotation and scale invariant BoW model was proposed for CT colonography [Aman et al., 2010]. But this invariance has been achieved by a feature descriptor that is used to extract features. It has used the SIFT, feature which has the property of invariant to rotation and scale.

Recently, another method has been proposed [Rahat et al., 2012] to incorpo-

rate spatial orientations of visual word pairs to improve retrieval performance. A spatial distribution of words is represented as a histogram of orientations of the segments formed by pairs of identical visual words (PIW). It has shown competitive performance in comparison with the SPM approach and other existing methods which incorporate spatial information. Even though this approach outperforms other existing methods, this method is only invariant to translation and scale but not to rotation.

5.5.3 Image Features

Local feature detectors and descriptors like Scale Invariant Feature Transform (SIFT), and Speeded Up Robust Features (SURF) which detect a number of interest points from an image, have been used in all the BoW approaches [Csurka et al., 2004, Yuan et al., 2011b] and the number of points may be a thousand or more than that. In contrast, the proposed RIBoW method uses global descriptors to generate BoW using sub-images which generate less data compared with local descriptors.

Different low-level features were used as a combination of colour, texture and shape as described in Section 5.2.1.

5.5.4 Vocabulary Generation

It was necessary to cluster the image features in this research. K-means is one of the simplest and the best-known unsupervised clustering algorithms that can be easily implemented for feature vocabulary generation. So visual vocabularies were generated by using the extracted features by feature descriptors from sub-images using K-means. Independent visual vocabularies were generated for each feature. As the main target was to achieve rotation invariance, circular image decomposition method, which we proposed in [Chathurani et al., 2015a] (Section 5.2.3), was used. Each image was partitioned in to eight sub-images as shown in figure 5.17. Then features were extracted from each sub-image using seven features. Firstly, features were extracted from the image database using this image decomposition, and a feature database was generated. Then seven independent visual vocabularies were generated. Here, the size of the visual vocabulary was 20 for Wang dataset and 40 for Caltech dataset and it is smaller compared with other methods which

have used the interest points [Aman et al., 2010, arszalek and Schmid, 2006, Yang et al., 2009].

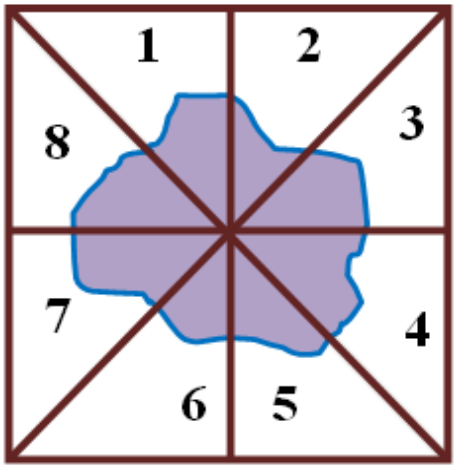


Figure 5.17: Circular image decomposition method.

5.5.5 Image Representation

The target of this proposed method was to achieve rotation invariance. So the BoW representation is different from typical representation. In a typical representation the order of sub-images is not taken into consideration, while in spatial representation it is. But spatial representation cannot handle the rotation invariance. This research work proposed a shifting operation to achieve rotation invariance. As shown in figure 5.18, figure 5.18.b is a rotated image of figure 5.18.a and figure 5.18.a is the un-modified image. When considering the distance, the measures of a.1 and b.1 histograms are different but the proposed operation has achieved the same histogram for both the image as shown in a.2 and b.2. As shown in figure 5.18, it has achieved the property of rotation invariance.

The process of generating RIBoW representation is as given below:

Step 01: N features from all the sub-images in the datasets were extracted by decomposing images, as shown in figure 5.17. N is the number of features used.

Step 02: The extracted data of each feature was clustered using K-means. Independent vocabularies were generated (codebook size is represented

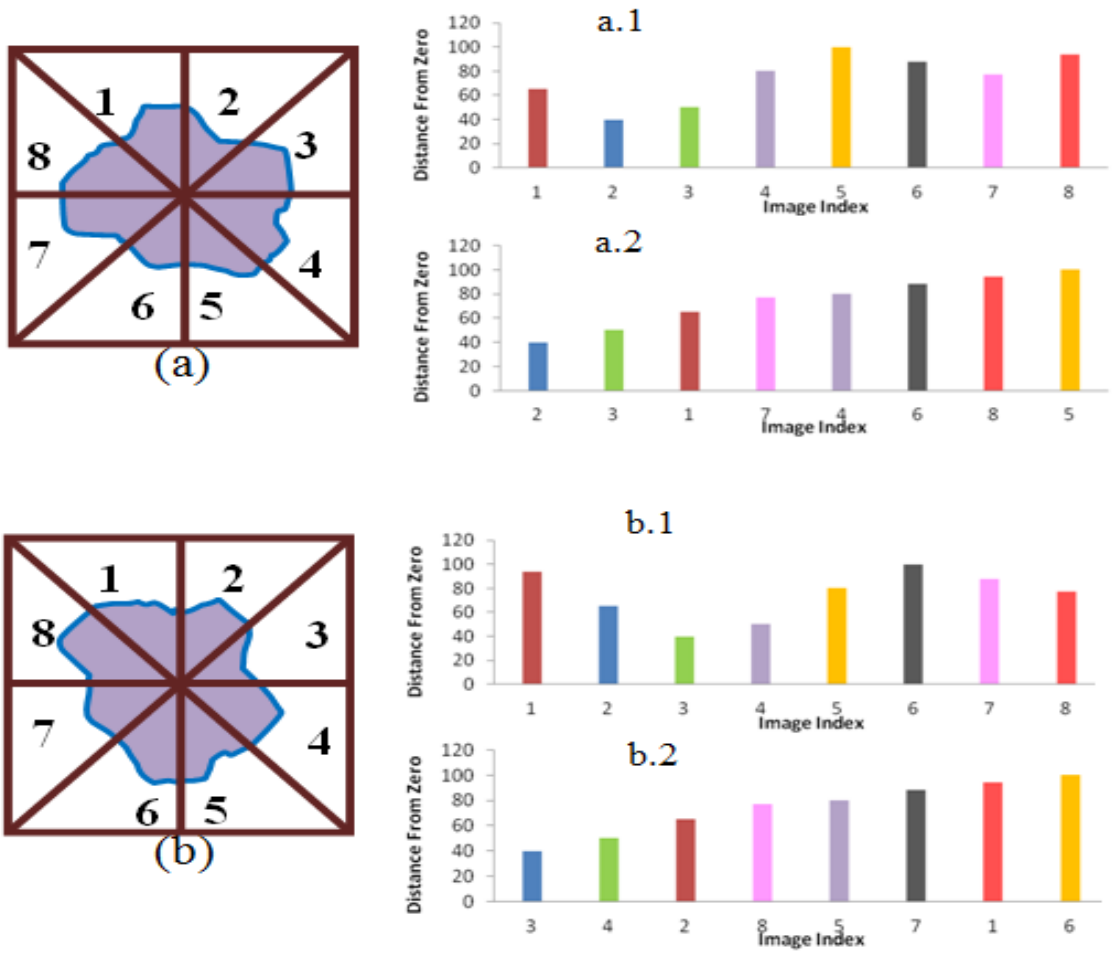


Figure 5.18: The new BoW representation method against the typical BoW representation method on rotation variance of an image. (a) the sample unmodified image, (b) the rotated sample image, (a.1 and b.1) the typical BoW representation of an unmodified image and the rotated image (the distance is measured from zero to the sub-image), (a.2 and b.2) the proposed representation of an unmodified image and the rotated image (ordered according to the distance from zero to the sub-image).

by K).

Step 03: Feature f_i from each sub-image of an image was extracted to represent each image using BoW representation (here f_i is used to represent features and i is the feature index where $i = 1, 2, ..., N$). While

doing so, the distance was found from zero vector to the feature vector, as shown in equation 5.3.

$$dist_{i\_k} = dist(f_i, 0) \tag{5.3}$$

Here $dist_{i\_k}$ is the distance from zero to feature f_i of sub-image k , where k is the sub-image index and $k = 1, 2, ..M$, M is the number of sub-images in an image;

Step 04: If the $dist_{i\_k} \leq dist_{i\_k+1}$ shifted the f_i up.

Step 05: Processed step 03 and 04 until $k = M$. Then the sub-images were in an order ($dist_{i\_k}$ ascending) of the full image.

Step 06: The nearest cluster centre to each sub-image was found in the given feature using a nearest neighbour search, then the cluster centre was assigned.

Step 07: All the images in the database are processed for feature f_i. using step 03 to 06.

Step 08: The above process was applied to all the features and the final BoW representation was generated. Each sub-image was represented using a feature index and cluster index which we proposed in Section 5.2.3 [Chathurani et al., 2014] as shown in equations 5.1 and 5.2.

Step 09: After the generation of BoW representation, signatures were generated for each sub-image, and then the generated signatures were used to generate a signature for the full image.

The order of sub-images in each image was taken into account when generating the final full image signature, as we required to achieve the property of rotation invariance.

5.5.6 Indexing and Searching

The final outcome of this method were binary image signatures and we kept them as a list, for sequential search. Hamming distance was used as a distance measure to find the similarity between images. The motivation to use signatures in representing images comes from the fact that computation time quickly becomes

a bottleneck when dealing with large databases and signature search engines can retrieve results from web-scale collections in milliseconds. So this representation can be used for large scale datasets.

5.5.7 Experimental Results

To evaluate the effectiveness of the proposed RIBoW approach, experiments were performed on the general purpose Wang dataset. As we were targeting specifically object retrieval, the large Caltech 256 object dataset was used to validate the system further. Detail of these datasets can be found at Section 3.1.1 and Section 3.1.3.

Experimental setup

All the compared BoW representations were implemented. We used the same codebook size for all the BoW representations with size K = 20 and K = 40, and the same features were used for evaluation. A standard histogram was generated by a 4 by 4 image decomposition. A three level pyramid was used (1 * 1, 2 * 2, 4 * 4) for SPM. Euclidean distance was used as the similarity metric for Histogram and SPM approaches, and Hamming distance was used as the similarity metric for signature-based representation. It must be noted that searching speed was much faster (in millisecond scale) than other compared approaches. Feature extraction time was nearly similar for all the cases as the same features and the same number of sub-images were used.

The most common evaluation measure in information retrieval is precision and it was calculated using equation 3.2. The proposed RIBoW approach was compared based on average precision by evaluating the top 20 retrieval results. A retrieved image was considered a correct match if and only if it was in the same category as the query image.

Results

Figure 5.19 shows the comparison of retrieval performance (AP@20) for different BoW approaches on the Wang dataset. It has shown that the simple standard histogram based representation has the worst performance (AP - 0.62). SPM BoW representations scores next (AP - 0.66) and RIBoW (Invariant Circular) achieves the highest (AP - 0.73). From these results it can be concluded that RIBoW

representation is superior to the typical histogram-based method and SPM. This can be extended for very large datasets by using local feature descriptors as they provide a large amount of data points per image.

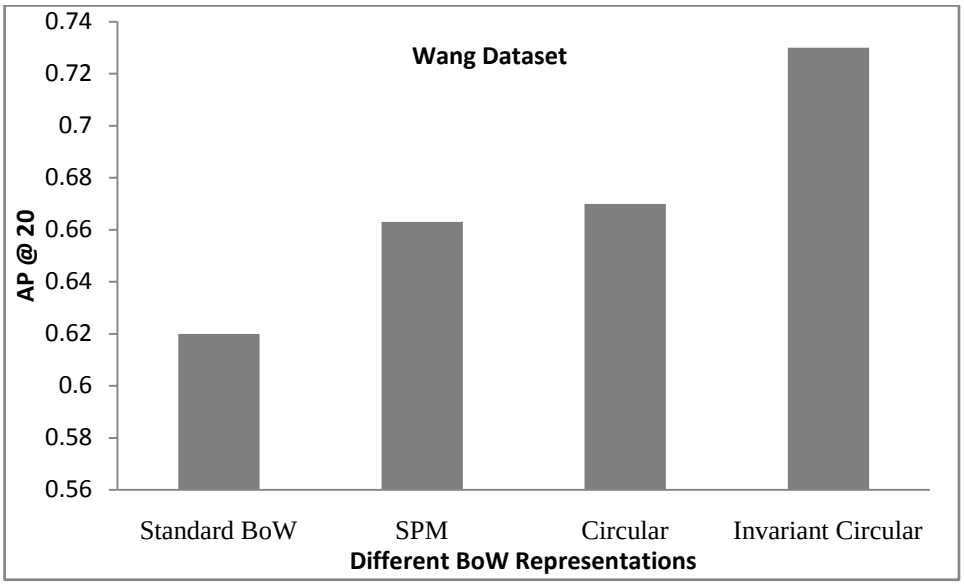
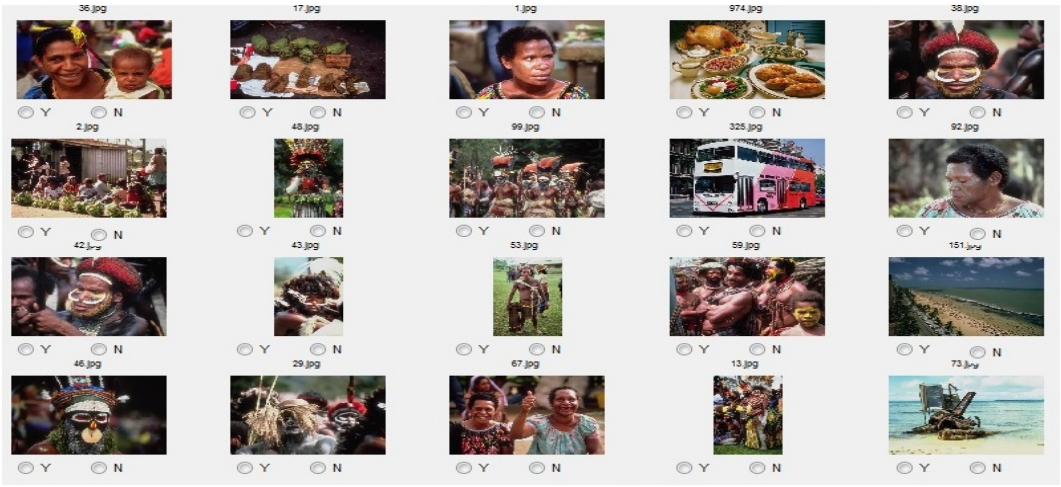


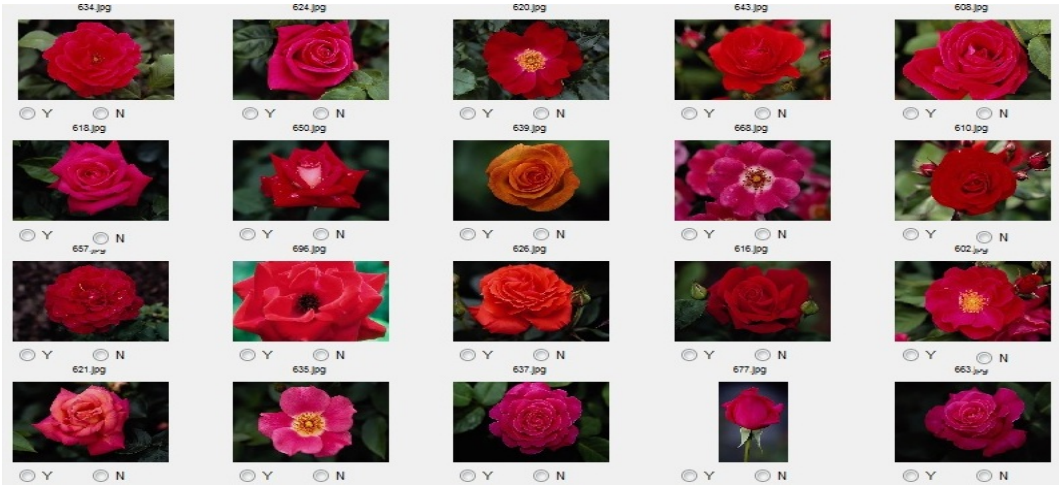
Figure 5.19: Performance comparison with different BoW approaches with RIBoW on the Wang dataset (AP@20).

Figure 5.20 shows the top 20 search results for some queries on the Wang dataset. In the top left is the query image. From that we can see that the performance is much higher for images with objects (figure 5.20.b, figure 5.20.c, figure5.20.e, figure 5.20.f - 20 out of 20) than scenery and cluttered images (figure 5.20.a - 16 out of 20, figure 5.20.d - 15 out of 20). When comparing results it is seen that the performance is higher for classes of images with objects than scenery. Therefore, this method will work very well for object-based image retrieval as we assume all the objects are towards the centre of an image.

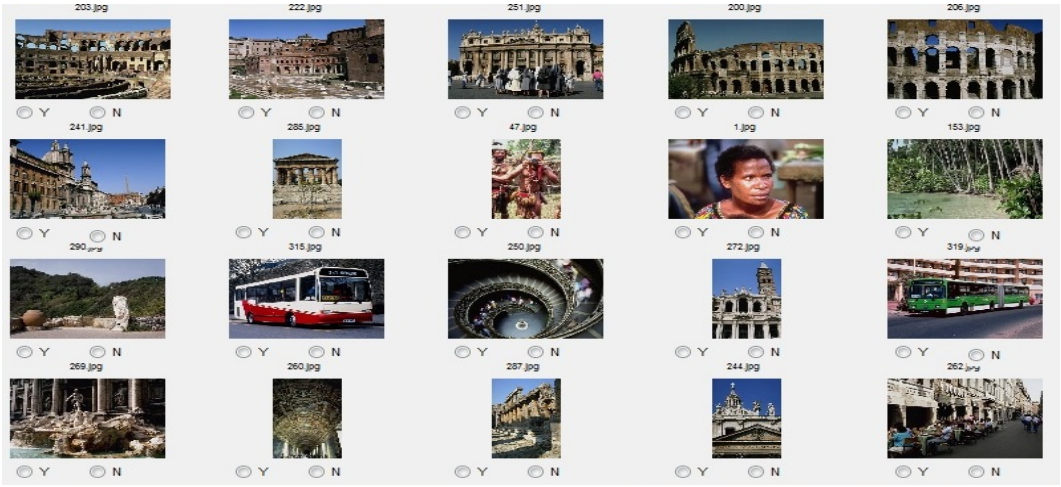


(a) Buses

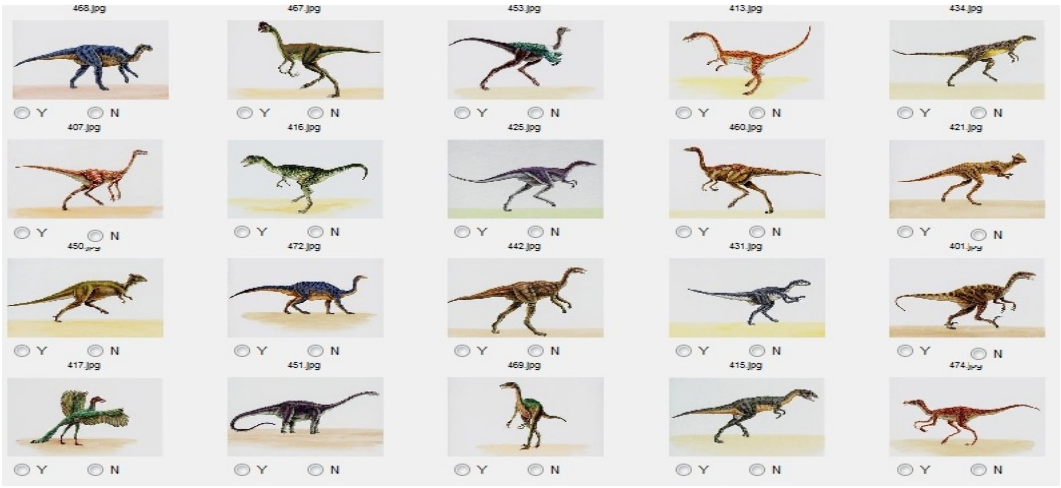
(b) Buildings



(c) Highway



(a) Beach



(b) Dinosaurs

(c) Forest

Figure 5.20: Search results of the RIBoW system for some queries on the Wang dataset (the query image is the top left most one). We can see that the performance is much higher for images with objects (figure b, c, e, f - 20 out of 20) than scenery, and cluttered images.

Figure 5.21 shows the comparison of AP @ 20 on Caltech 256 dataset. It shows that the simple standard histogram-based representation has the worst performance (AP - 0.085). SPM BoW representations scores next (AP - 0.113) while RIBoW (Invariant Circular) achieves highest (AP - 0.147). Even though the performance for this dataset is 0.147, it is substantially higher compared with other approaches. When we considered the performance of each class separately, some classes had higher performances (ex: less clutter, bigger objects) while some had lower (ex: cluttered, small objects with different backgrounds).

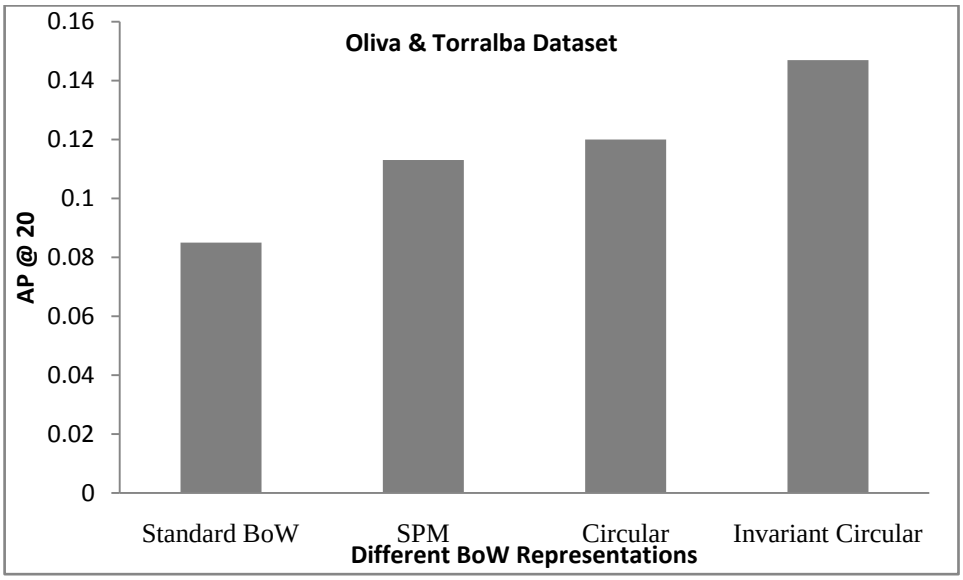


Figure 5.21: Performance comparison with different BoW approaches with RIBoW on the Caltech dataset (AP@20).

Figure 5.22 shows the results of some queries on Caltech 256 dataset which had higher retrieval performances. From that, it is seen that the proposed RIBoW approach has good potential to retrieve similar images. Except for the first row, all the images are relevant to the query image in the figure. Figure 5.23 shows an example of the retrieval results of a query for different BoW approaches. Relevant images are clicked as Y (yes) under the image. This figure further illustrates that the proposed RIBoW approach is much better when compared with BoW approaches. Circular representation provides correct results in five out of five top-ranked images but invariant circular representation (RIBoW) provides more

semantically similar images.

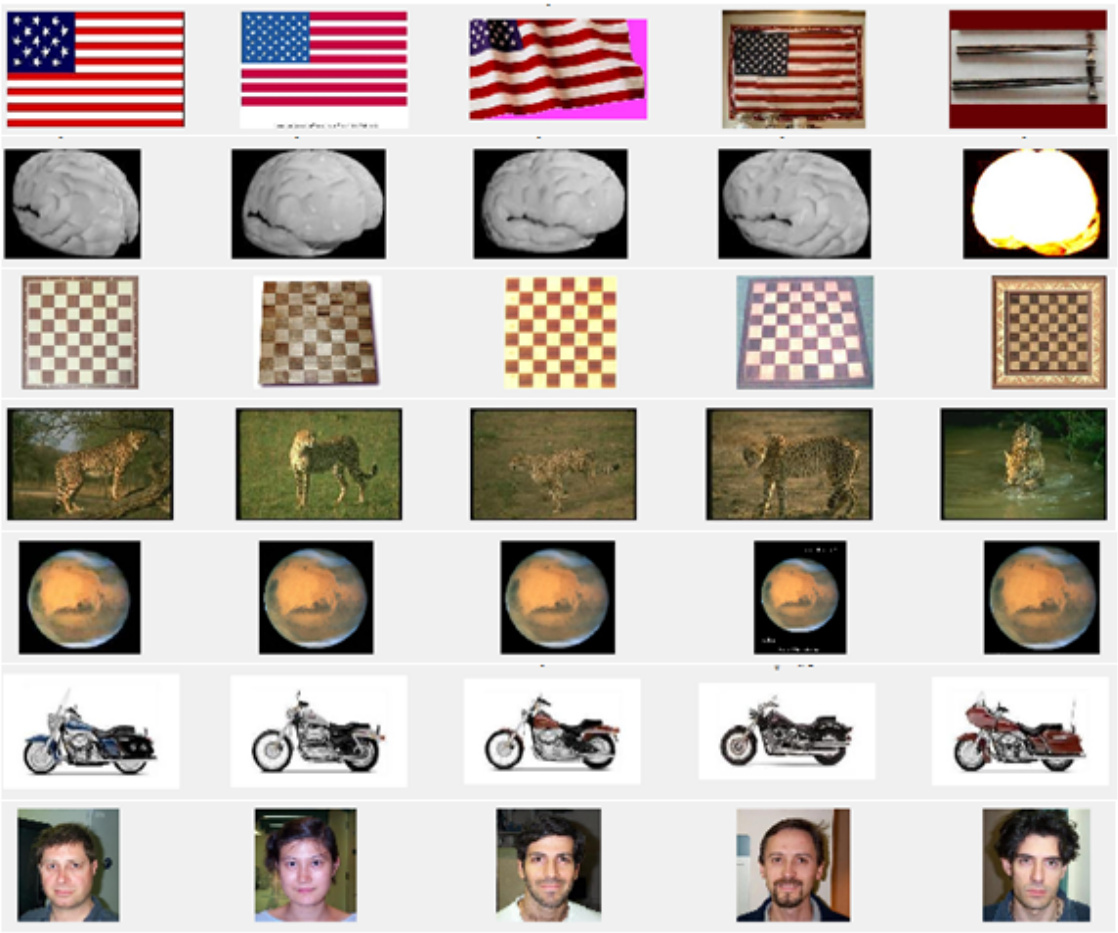


Figure 5.22: Search results of the RIBoW system for some queries (some classes which achieved higher retrieval performance) on the Caltech 256 dataset (query image is the first one).

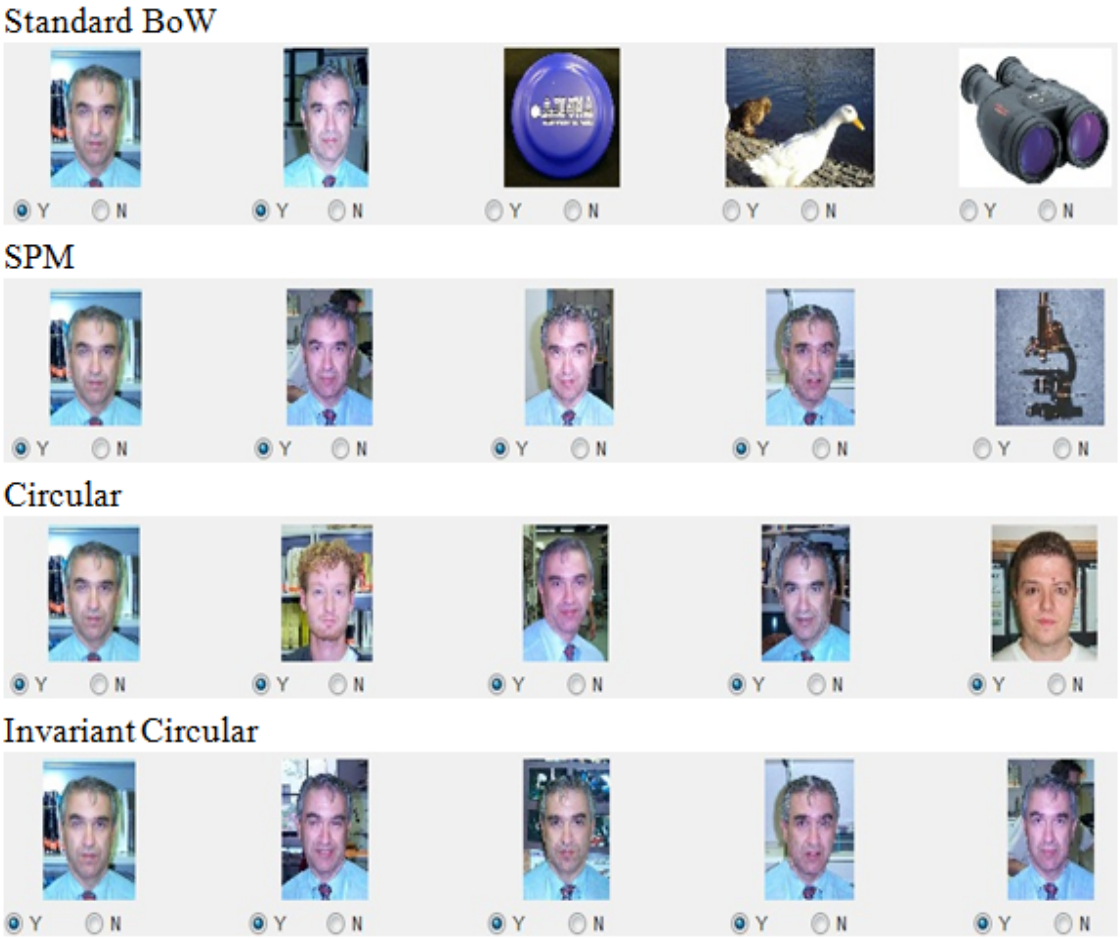


Figure 5.23: Search results for a query on the Caltech 256 dataset. Each row shows the top-ranked outputs produced by different BoW approaches, and the relevant ones are ticked.

5.5.8 Section Summary and Conclusions

A simple rotational invariant bag of visual feature approach is proposed based on circular image decomposition and binary signatures. It was found that the proposed approach has good potential to retrieve correct images, especially objects. The proposed approach was validated using two general datasets. Unlike other BoW representations, this approach can be extended to large datasets, as this can handle scalability issues due to signature-based representation. The proposed

approach showed superior performance in retrieval quality relative to standard histogram, SPM and circular invariant representation.

5.6 Chapter Summary and Conclusions

This research initially found the most suitable building block to extract features and then proposed another image decomposition method which was used for rotational invariant object retrieval by outperforming typical representations later. The most suitable image feature combination was found using leave-one-out feature selection. Sub-image is the main building block of this research and hence it was characterised by colour, texture and shape features using BoW with the feature index and the index of the nearest cluster centre. Then RI was introduced to CBIR by applying RI on the generated file of symbolic representation. Retrieval quality was depended on signature size and the one factor that influenced the retrieval speed also was signature size, which will be discussed later in this dissertation. 8192 bits signature size showed better retrieval performance while it balanced the trade-off. Therefore, an 8192 bits signature size was used in the system. Then the system was empirically evaluated to compare with existing systems and it outperformed them in retrieval quality. Finally, a rotation invariant bag of word approach was proposed using the binary image signatures and was evaluated on several datasets to show the effectiveness of the proposed approach.

After all the experiment, we wanted improve the retrieval quality of the system by reducing semantic gap by applying relevance feedback to the existing signature-based system. Then we could be able to propose a rank based pseudo relevance feedback scheme which will be introduced in Chapter 6. First pass retrieval results from this chapter were used to evaluate the proposed relevance feedback approach.

Chapter 6

CBIR with Pseudo Relevance Feedback

Chapter Organisation

The main objective of this chapter is to show the application of Relevance Feedback (RF) approach into the Image SIGnature-based, Content-Based Image Retrieval (CBIR-ISIG) system to improve retrieval performance. A simple introduction about RF and brief overview of the CBIR-ISIG system is given in the Section 6.1 and 6.2 respectively. Details about the design and implementation of the proposed Rank-Based Pseudo Relevance Feedback (RB-PRF) approach is provided in Section 6.3. Performance evaluation of the system, parameter settings and analysis of results are provided in Section 6.4. The chapter summary and conclusions are included in Section 6.5. The original contributions discussed in this chapter resulted in publication (v) in the List of Publications. The RB-PRF approach was used here to improve the retrieval quality obtained using the binary image signatures, proposed in Chapter 5.

6.1 Introduction

Relevance feedback refers to the online processing algorithm which interacts with the user of the search engine. The use of RF requires a user to provide the feedback (FB). This requirement has the following implications for situations in which RF

can be used:

- An increase in the amount of effort required from the user in order to interact with the search engine; user time demands per iteration are high.
- Systems that make use of RF can only be used in the context of interactive queries and cannot be used for automatic processing.

This precludes the use of, and limits the effectiveness of, using actual RF in many situations. However, even in cases where explicit user feedback is not available, some aspects of the RF approach can be used by making the often reasonable assumption that the first pass results returned by the search engine will include a sufficiently large number of images that match the user's information need. The system then makes use of that set of top-ranked initial results as if it were actual positive feedback provided by the user. This technique is often referred to as Pseudo-Relevance Feedback (PRF). The assumption is that despite being noisy feedback, it will still help (promote) pick up missed relevant documents in the initial retrieval list.

Some techniques used in CBIR for RF include query point movement [Cui and Zhang, 2007, Su et al., 2011a], re-weighting feature vector [Jing et al., 2002, Su et al., 2011a], support vector machine [Wu and Yap, 2006, yi Lee and Lee, 2013, Yap and Wu, 2007], neural networks [Wang et al., 2006, Ko and Byun, 2002] and statistical learning [Rahman et al., 2007]. Among these approaches, the use of query point movement and feature vector re-weighting are the most widely used methods in RF. This latter method shares similarities with the proposed RB-PRF approach. Feature vector re-weighting techniques update the weights of feature vectors using heuristics so as to emphasise the components that are most shared among relevant images. This may help to retrieve relevant images, while de-emphasising those components which appear more commonly in irrelevant images. Unlike the proposed RB-PRF method, most feature vector re-weighting techniques consider real-valued vectors rather than binary signatures, as shown in this thesis, and they further require access to collection statistics to perform the query update.

Motivated by the concept of PRF, the RB-PRF approach is proposed to improve retrieval performance by applying PRF on binary image signatures. PRF is not a new concept. However, two new contributions are made to the traditional

method - the first is the original use of document signatures directly in feedback processing, as opposed to traditional approaches which return to the original images (or documents). The second is the incorporation of the rank order of the initial results in utilising feedback to re-rank the results rather than assuming equal importance, which is performed in the traditional methods. In addition, this provides faster retrieval performance in feedback by considering only a subset of the full dataset for comparison.

Image representation is derived through sub-image decomposition, and a full image signature is generated by the sub-image signatures. Image signatures are fixed length binary strings derived through a form of locality sensitive hashing [Geva and De Vries, 2011]. The derivation of an image signature from an image feature is described in Chapter 5. Sub image signatures are generated from low-level features of each sub-image.

A retrieved image list of length K, (much smaller than the collection size) which is small enough to be practical, is re-ranked through RB-PRF by using the top N image signatures as relevant image signature examples and the bottom N as non-relevant examples, where $N << K$. K is selected to be large enough that it can be assumed that the bottom N results are unlikely to be relevant, yet small enough to ensure that the bottom N results are still similar to the query signature, albeit irrelevant. The entire feedback-based re-ranking is performed in signature space. Returning to the original image representation, as done in conventional PRF, is not necessary here. Here we re-rank only the retrieved list of first K binary image vectors against the generated feedback signature. Results presented here show that RB-PRF achieves effective and efficient RF in CBIR and a considerable improvement in retrieval performance over earlier approaches.

The RB-PRF technique deviates from traditional CBIR approaches as it works directly with image signatures, rather than image features. Signatures are locality-sensitive hashes (binary strings of fixed length) which are used to represent images for the purpose of searching [Faloutsos and Christodoulakis, 1984, Chappell and Geva, 2015]. Furthermore, in our implementation of the RF mechanism, the top-ranked signatures are already resident in memory and thus there is no need to work with the original documents at run time. The use of signatures and the allocation of resident memory for their storage allow for an extremely efficient

retrieval method, both in terms of memory usage and run-time. In addition, RB-PRF explicitly incorporates the original rank of the results used as implicit relevance indicators, thus deviating from the common use of PRF that ignores the rank positions of the feedback documents.

As PRF is an unsupervised learning process, it cannot be guaranteed that the top-ranked images are all relevant. If the initial CBIR system output is better, the re-ranked results with PRF are usually better still; conversely, if the CBIR system provides poor initial results, the re-ranked results with PRF could lead to worse results because of the noisy/off-topic image feedback. The CBIR-ISIG system which was proposed in Chapter 5, shows significant performance, suggesting that searching the initially retrieved list using RB-PRF can lead to potential improvement in performance.

Extensive experiments have been carried out to study the behaviour and optimal parameter settings of this approach. Empirical evaluations based on standard benchmarks (Wang, Oliva and Torralba, and Corel datasets) demonstrate the effectiveness of the proposed approach in improving the performance of CBIR in terms of recall and precision. Before going further into the detail description about the feedback approach, a brief overview of the CBIR-ISIG system is given in Section 6.2.

6.2 First Pass Retrieval

Image signatures are used in the retrieval process and it is found that signatures of 4K-8K bits in size are sufficient to achieve effective improvements over the baseline retrieval method, while substantially reducing the run-time of the retrieval process which is described in Chapter 5. Furthermore, according to figure 5.9 in Chapter 5, it can be seen that larger size signatures results in diminished performance quality and an 8K bit signature size is optimal, as it a is good compromise between signature size and quality. The longer the signature size, the slower the searching process. The Hamming distance is used to search images and it is effective since it can be performed with compact low level machine instructions. A full description of image signature generation is provided in Chapter 5 and brief overview of the system can be taken from figure 6.1.

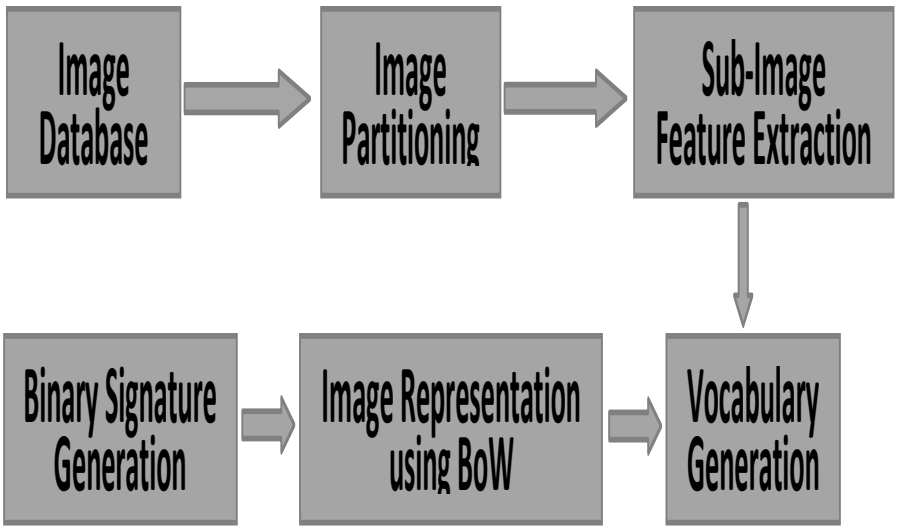


Figure 6.1: Process of binary image signature generation.

6.3 Pseudo Relevance Feedback Approach

At the query time in RB-PRF the search initially produces a list of images with signatures ranked from the most similar to the query signature, to the least similar. The top-ranked K signatures are then used in RB-PRF process. RB-PRF takes the resulting list L, $L = \{I_i\}_{i=1}^{K}$, where K is the size of retrieval list, from the initial search. Then binary vectors of the top N (the size of the PRF list) signatures and bottom N from the initial result list L (K binary vectors) are considered as relevant and irrelevant respectively, to generate a feedback signature. Each selected signature is represented by values of $\{-1, 1\}$ with 1 - bits being interpreted as 1 and 0 - bits interpreted as -1 as described in the Section 5.2.4. Theoretically, they are represented as sequences of 1 and -1, but computationally ,they are stored as sequences of 0-1 bits. Binary signatures are transformed into real valued vectors as a weighted linear combination of the feedback signatures, weighted by rank, with negative feedback ranked in reverse order. Each vector is multiplied by the scaling factor S, such that the signatures that are closer to the top of the list contribute more to the new real valued signature. The scaling factor S is defined as in equation 6.1

$$S = e^{-(i-1)*(\frac{i}{W*N})} \tag{6.1}$$

Where N is the number of signatures being sampled and i is the rank of the signature in the initial results list ($i = 0, ..., N-1$). The term w is a decay factor determined empirically; the value 3 of S was found to be best in our experiments (see figure 6.7 in Section 6.4). The results are not very sensitive to the value of w, which determines how fast the feedback from images decays with rank and this value works well over a wide range of collections and experiments. The feedback vectors are then added together, independently and specifically, two separate feedback vectors are generated - one from the pseudo-relevant signatures and one from the pseudo-irrelevant signatures. The vector generated from the irrelevant signatures is then subtracted from the vector generated by the relevant signatures. Finally, the vector is "squashed" back to a binary representation $\{1, 0\}$ simply by taking the sign bit. The resulting binary signature is the feedback signature. The signatures in the result list that are to be re-ranked are then sorted according to the Hamming distance from the new feedback signature. Note that the entire re-ranking process takes place in signature space without ever going back to the original image features. Furthermore, the initial list of signatures that are being re-ranked is already in memory following the initial search process, so the process is computationally efficient.

This method, which is different from classic Rocchio's algorithm, is concerned with identifying terms that have significant correlation between relevant results and irrelevant results. It adds search terms from the relevant documents, ostensibly giving greater weight to documents that share similarities with the set of relevant documents and lesser weight to documents that share similarities with the set of irrelevant documents [Rocchio, 1971, Ishikawa et al., 1998, Lu et al., 2000, Shaw, 1995]. A full document level representation is used in the new RF approach and it does not require to go back to the original document to collect term statistics, which is much simpler than the Rocchio algorithm. Furthermore, in our algorithm the top-ranked signatures are already memory resident and there is no need to work with the original document on the fly, which keeps our algorithm computationally efficient.

Figure 6.2 provides a detailed description of the RB-PRF process. Suppose that the PRF list size is 5 ($N = 5$) and further suppose that signature size is 4-bit in PRF approach as shown in figure 6.2.a. These are converted to real valued

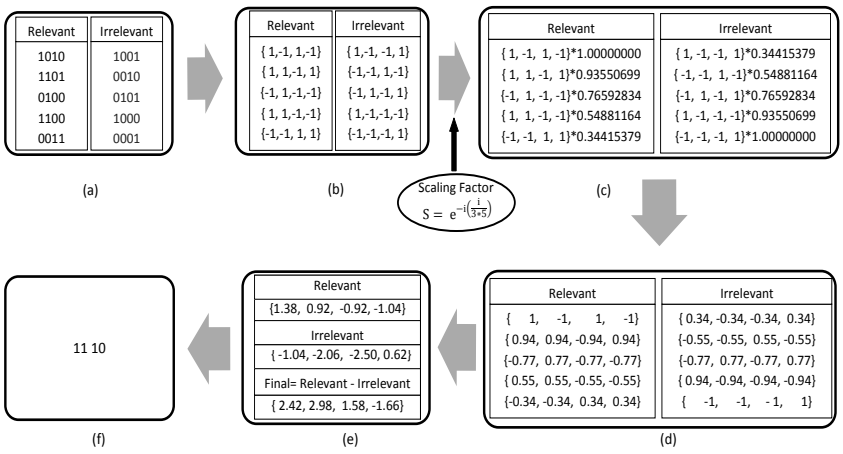


Figure 6.2: Toy example to show the process of PRF using a toy dataset with signature size four bits and sample size (N) five (which are considered as relevant).

vectors as shown in figure 6.2.b, and multiplied by the scaling factor S using equation 6.1 ($N = 5$ in this example and i is from 0 to 4 in these signatures). Figure 6.2.c shows the process of scaling and figure 6.2.d shows the results. The scaling formula used here ensures a smooth decay of feedback with increased rank position and was chosen after experiments with cross-validation. Scaling performs significantly better compared with the method without scaling. Vectors are added and subtracted, as shown in figure 6.2.e. Finally, a binary image signature is generated, as shown in figure 6.2.f. This is then used to re-rank the initial search results. If you need further details to understand how the scaling factors varies with w and i, refer Figure 7.8, which shows values of scaling factor (S) with variation of w as a function of i.

6.4 Evaluation of the Pseudo Relevance Feedback Approach

Different experiments were carried out on the proposed RB-PRF to study its behaviour and effectiveness. All the experiments, evaluation methodologies and results are described in this Section.

6.4.1 Evaluation Measures

The RB-PRF approach was evaluated on different datasets using different evaluation measures to study the effectiveness of the approach in CBIR by considering retrieval quality. Experiments were performed on several general purpose image datasets Wang, Oliva and Torralba and corel. The details of these datasets can be found at in Section 3.1.1, Section 3.1.2 and Section 3.1.5.

The evaluation measures used for the evaluation of the system are precision, recall and precision at n using equations 3.2, 3.1, 3.3 and 3.4.

Three evaluation methods of full ranking, freezing and residual ranking were used to empirically evaluate the proposed RB-PRF approach as described in Section 3.3.

AP@n was calculated for the Wang and Oliva and Torralba datasets. AP@20, AP@50 and AP@100 were calculated for each class to compare with existing systems. RB-PRF was compared with the baseline signature-based system (CBIR-ISIG), and with existing systems from the literature. The system used for evaluation is shown in figure 6.3. It may be noted that we do have proxy for both positive feedback and negative feedback in our approach, as mentioned in Section 6.3.

The RB-PRF system was evaluated with different settings to study how they affect its behavior, including the sampling and re-ranking of different portions of the result lists. 8K bit signatures were used, as these were found to produce a good trade-off between efficiency and precision.

6.4.2 Evaluation Methodology and Results

First, we explored the role of scaling factor (S) and (w) values on the effectiveness of our RB-PRF. This enabled us to set effective parameters for use in the experiments. The scaling factor was used to generate the feedback signature from selected images by incorporating their rank in the list. Several empirical scaling functions were experimented with, all modelled as decaying functions of increased rank positions, hence attributing greater importance to top-ranked results. Figure 6.4, figure 6.5 and figure 6.6 show the MAP results obtained by varying scaling factor functions on the Wang, Oliva and Torralba, and Corel datasets (these results were obtained with signatures of 8K bits in size).

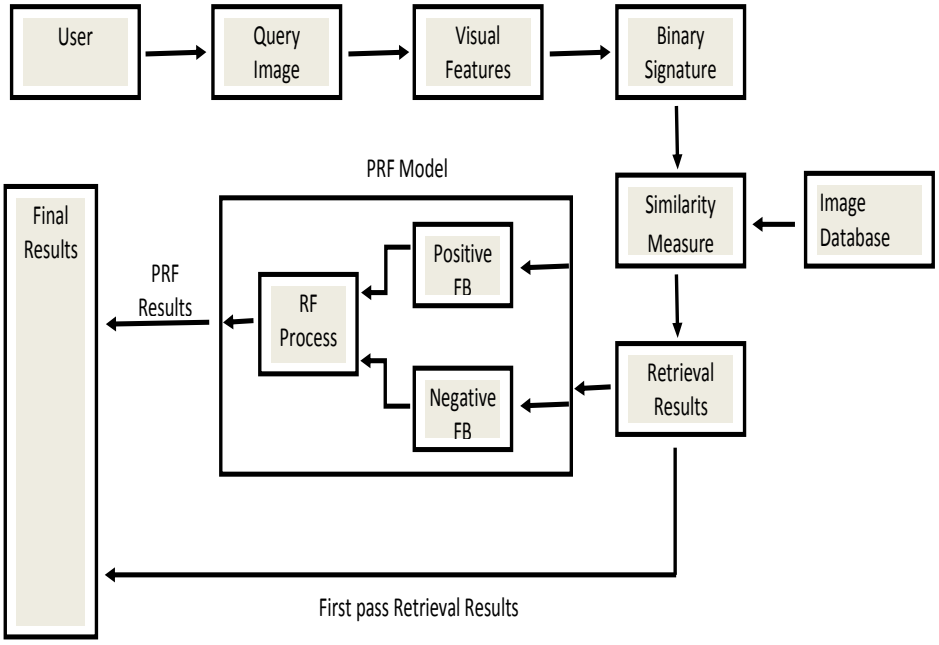


Figure 6.3: CBIR-ISIG system with PRF. Images are initially ranked using binary image signatures and then re-ranked with the PRF before the final results are shown to the user.

$1/(hamming\ distance)$, $1/sqrt(hamming\ distance)$ $1/exp(i)$ gives exactly the same results as the original query in the figures because $exp(-1)$ is greater than the sum of all the values from $exp(-1)$ to $exp(-\infty)$. Therefore, the first matching signature dominates the retrieval quality. Precision by applying scaling factors of $1/(hamming\ distance)$ and $1/sqrt(hamming)$ is also dominated by the first matching signature if the hamming distance is 0 for the first match. Moreover, if the difference between first match and the next is bigger, the first matching signature may overshadow every other feedback signature. $S = 1$ gives equal importance to all the feedback signatures and the result was drastically reduced when the feedback sample size was increased because all the images that were considered relevant may not have been relevant, according to the classification. Results for $1/sqrt(i)$ reduced more quickly when the feedback sample size was increased than $1/i$ as it gave undue importance to latter images in the feedback sample. According to these figures $e^{-(i-1)*(\frac{i}{3*FB\ \ sample\ \ size})}$ scaling function gave the best performance. There are various goals to be balanced in the scaling function.

First of all, we want the higher-ranked (lower i) results to contribute more to the feedback signature than the lower-ranked results, but not so much so that the lower-ranked results are entirely invalidated by the higher-ranked results. Next, we want the gap between ranks to increase with the increasing rank, so that it does not face the problem you see when using a linear function, with a lot of low-ranked results out-voting (or rather, adding noise to) a consensus among the high-ranked results. An exponential decay factor helped adjust that. Therefore, we proposed the scaling function with exponential decay for the RB-PRF system, as shown in Equation 6.1.

We experimented with the w in the scaling factor S to find the most suitable value for this RB-PRF. Figure 6.7 shows the average precision for the changing w value on both the Wang and Corel datasets. It demonstrates that $w = 3$ gives the best output. Therefore, $w = 3$ was used in the RB-PRF system. However, the results are not very sensitive to the value of w.

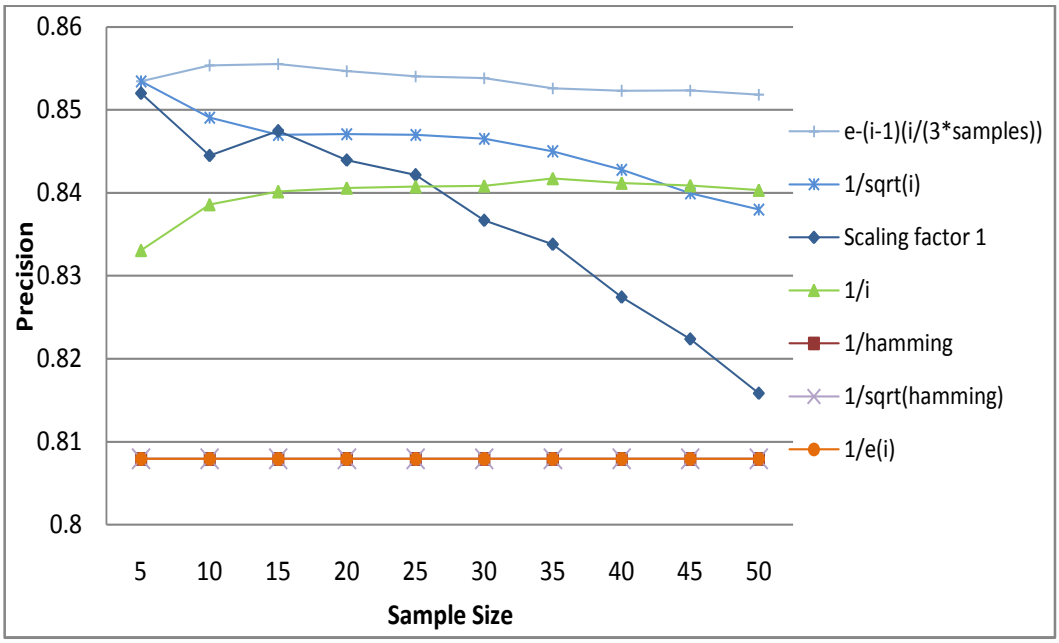


Figure 6.4: Precision vs scaling factor on the Wang dataset.

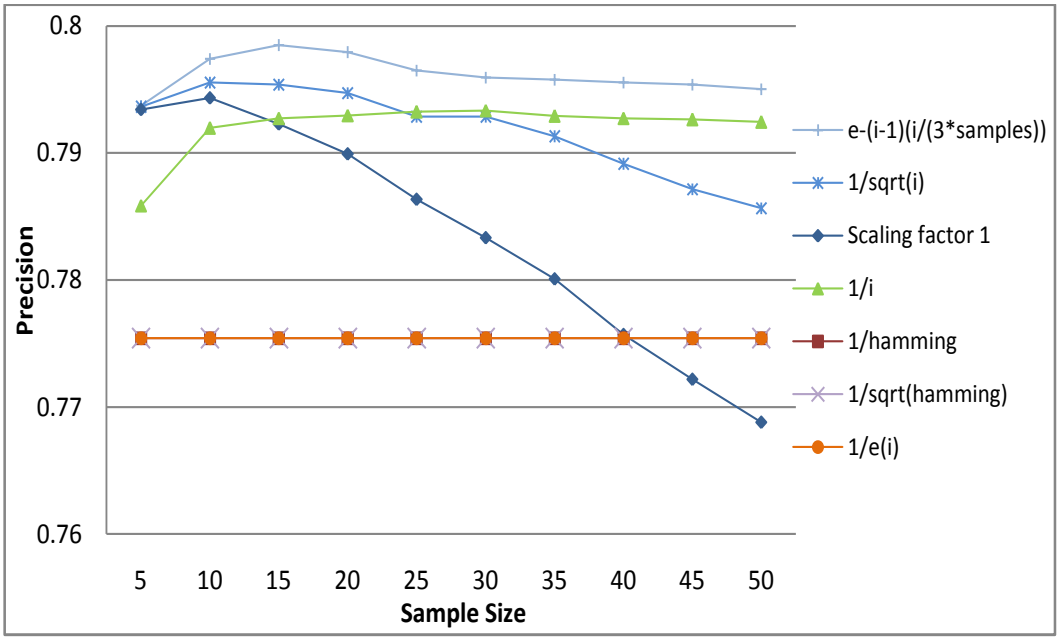


Figure 6.5: Precision vs scaling factor on the Oliva and Torralba dataset.

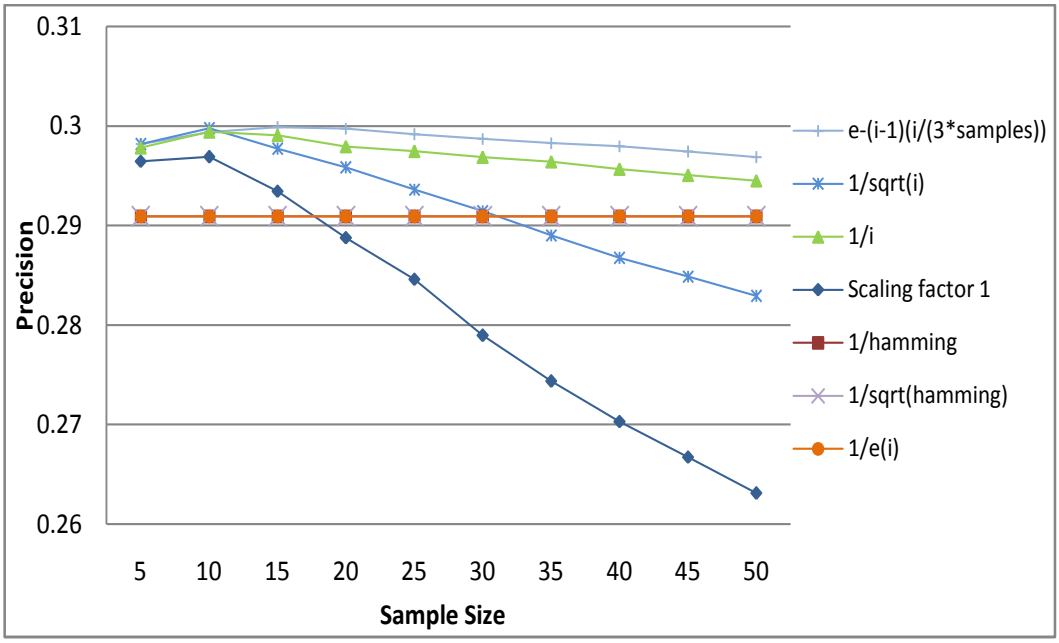


Figure 6.6: Precision vs scaling factor on the Corel dataset.

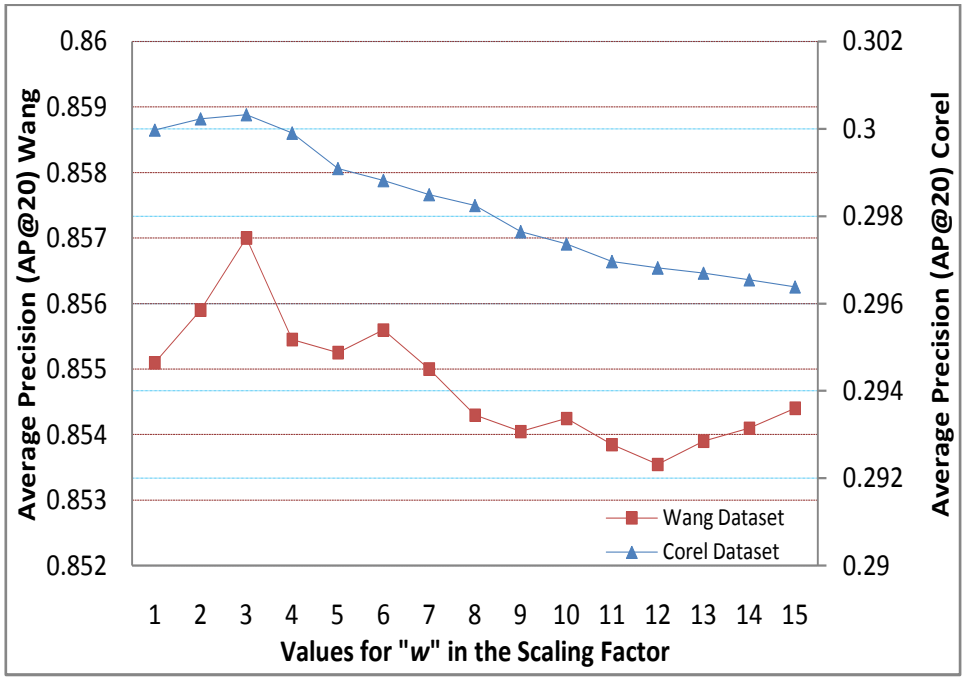
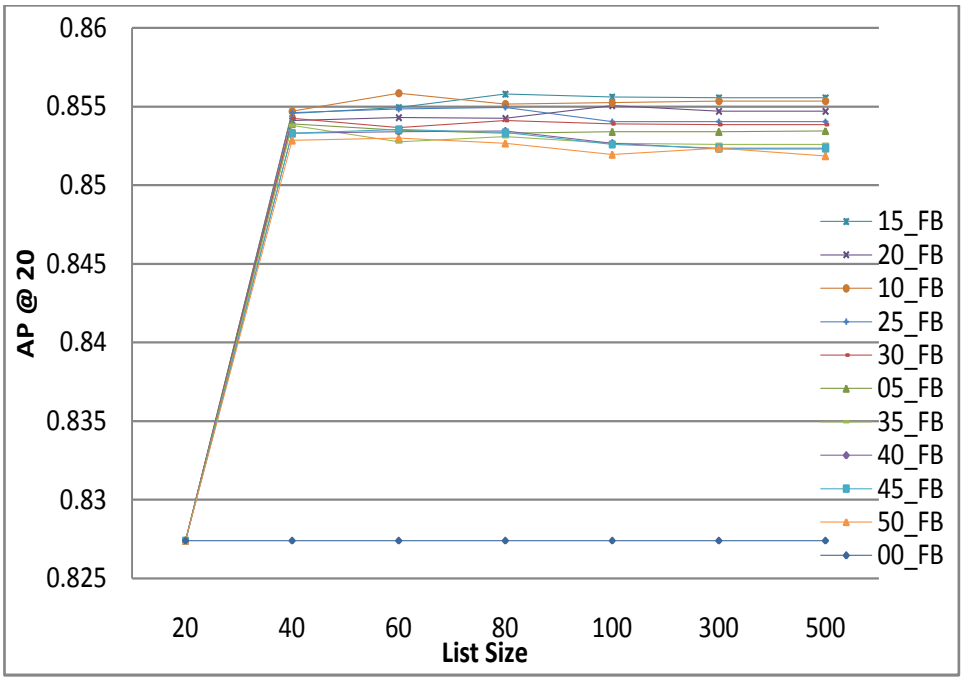
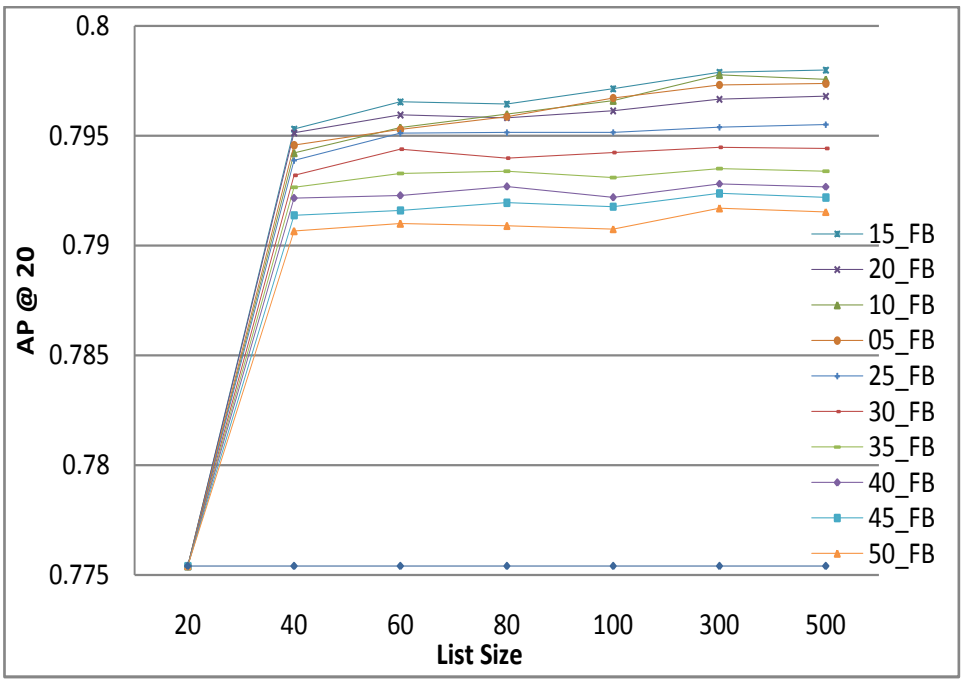


Figure 6.7: Performance variation with the change of w in the scaling factor(MAP@20) on the Wang and Corel datasets. Here feedback sample size is 20 with 8K signature size.

Secondly, the effectiveness of the RB-PRF method as a function of feedback sample size and the re-rank result list size was analysed. Figure 6.8.a and figure 6.8.b show the comparison of MAP and AP@20, and figures 6.9.a and figures 6.9.b show the comparison of MAP, AP@100, from no feedback to feedback size 50 on the Wang dataset and Oliva and Torralba dataset with the changing size of the re-rank list size from 20 to 500. Entries in the legends are listed from top to bottom according to the order of performance from best to worst in each figure. Figure 6.8.a shows a MAP increase of 3.0% from no feedback to PRF for the Wang dataset, and a MAP increase of 2.5% for the Oliva and Torralba dataset as shown in figure 6.8.b. The behaviour of the PRF on both the datasets are quite similar. The Wang dataset MAP peaks at 0.856 at top-15 feedback and the Oliva and Torralba dataset peaks at 0.798 at top-15. Figure 6.9.a shows a MAP increase of 7.5% from no feedback to PRF (0.694) on the Wang dataset and 4.5% for the Oliva and Torralba dataset (0.694), as shown in figure 6.9.b. These results demonstrate the effectiveness of the proposed RB-PRF in improving retrieval performance.

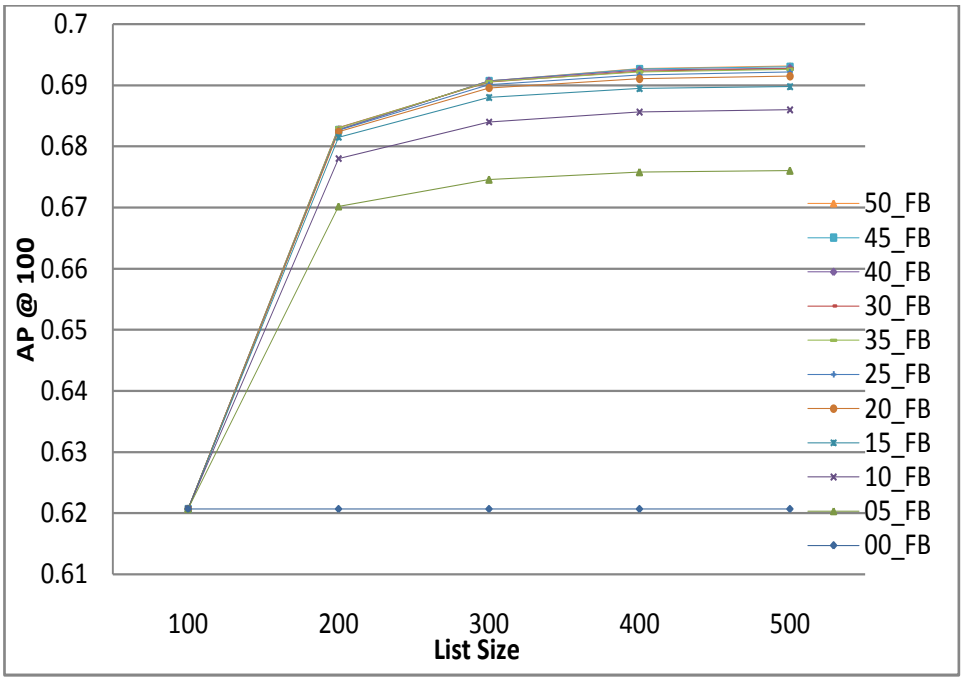


(a) Wang Dataset AP@20

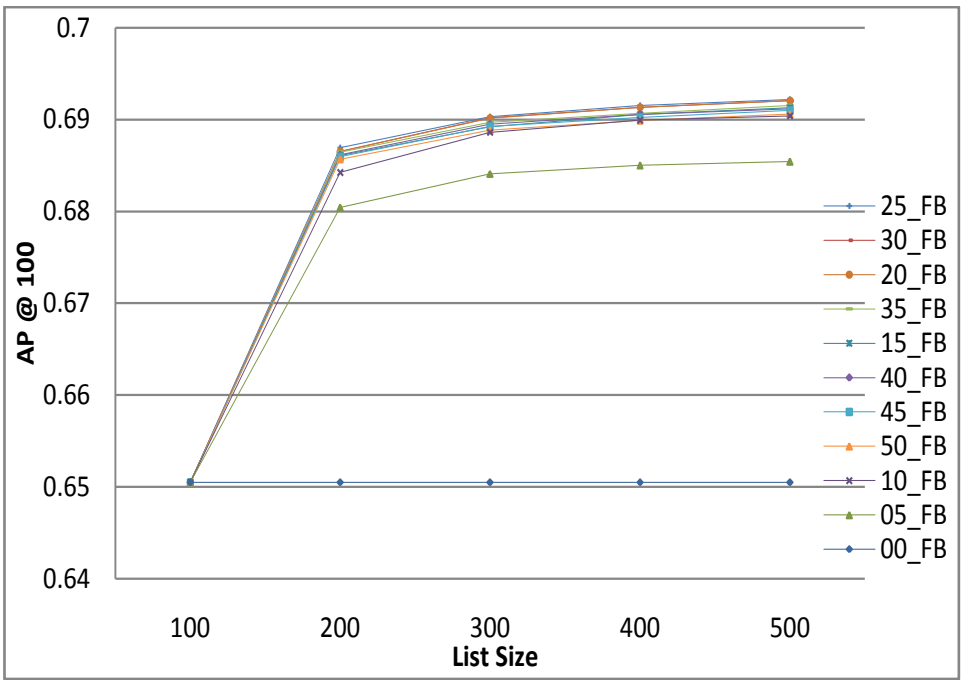


(b) Oliva Dataset AP@20

Figure 6.8: Comparison of AP@20 with the variation of the re-rank list size and sample size for RB-PRF with positive PRF. Each line corresponds to a sample size.



(a) Wang Dataset AP@100



(b) Oliva Dataset AP@100

Figure 6.9: Comparison of AP@100 with the variation of the re-rank list size and sample size for RB-PRF with positive PRF. Each line corresponds to a sample size.

Figure 6.8 and figure 6.9 show that the precision is increasing with the size of the re-rank list up to a certain depth and saturates after that. AP@20 is increased up to a list size of 80 and 300, and AP@100 is increased up to a list size of 500. Figure 6.8 shows that precision (AP@20) is increased with the size of the feedback sample, up to 15 and 20, and then starts to decrease with the increasing size of the sample. This is because of the increased noise level arising from the reduced precision of the feedback signatures list as it is lengthened.

We summarised the above mentioned data for both the Wang and Oliva and Torralba datasets in figure 6.10 and figure 6.11. The technique exhibited the same trend on both datasets and no significant improvement was found for feedback sizes above 20. Therefore, we used a feedback size of 20 in our experiments. According to figure 6.11, effectiveness appears to increase with the size of the list of images to re-rank, however, this plateaus after list sizes of 80 on the Wang dataset and 300 on the Oliva and Torralba dataset are reached (500 for both the datasets for AP@100). According to these experiments, it was found that a smaller feedback sample size and a larger re-ranked list size appeared to be the most suitable settings for the technique. However, the retrieval effectiveness was likely to improve if there were relevant images in the initial list which were ranked at a sufficiently high level, so they could be promoted in the re-ranked results list. Relevant results that were far too deep in the initial list were unlikely to be promoted and so the effectiveness improvements leveled off. Therefore, we used a re-rank list size of 500 in our experiments.

To further validate the system, the RB-PRF system was evaluated using different evaluation criteria. Table 6.1 shows AP@20 with the changing sample sizes for full ranking, freezing and residual ranking on the Wang dataset. The RB-PRF system shows its best performance at 15FB (sample size) and it has 79% average precision even with the residual ranking. Table 6.2 shows AP@n for same measures on the Wang dataset. The RB-PRF system shows more than 59% precision (AP@100) even with residual ranking (sample size $N = 15$).

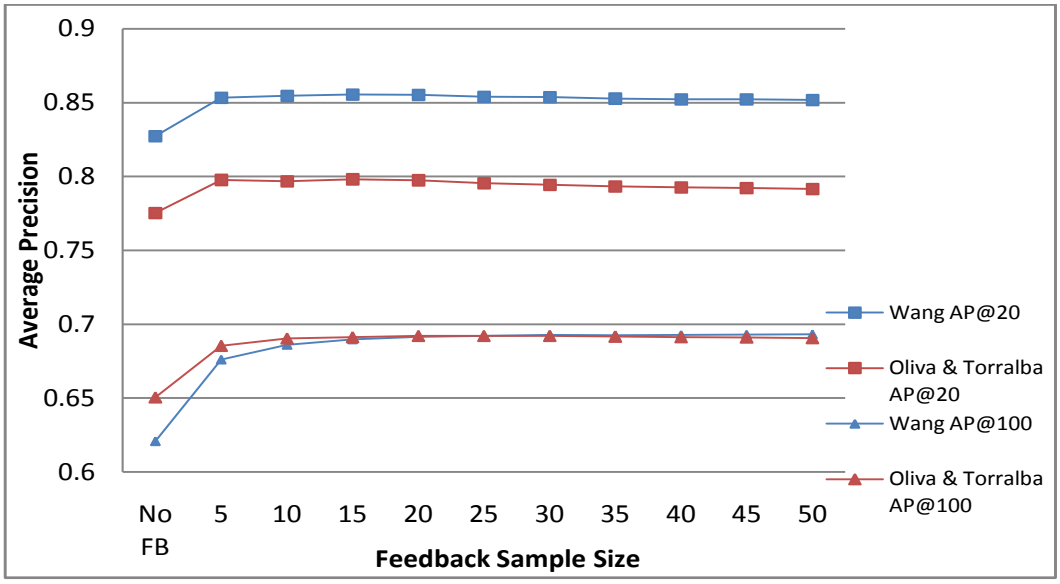


Figure 6.10: Performance variation with the change of RF size.

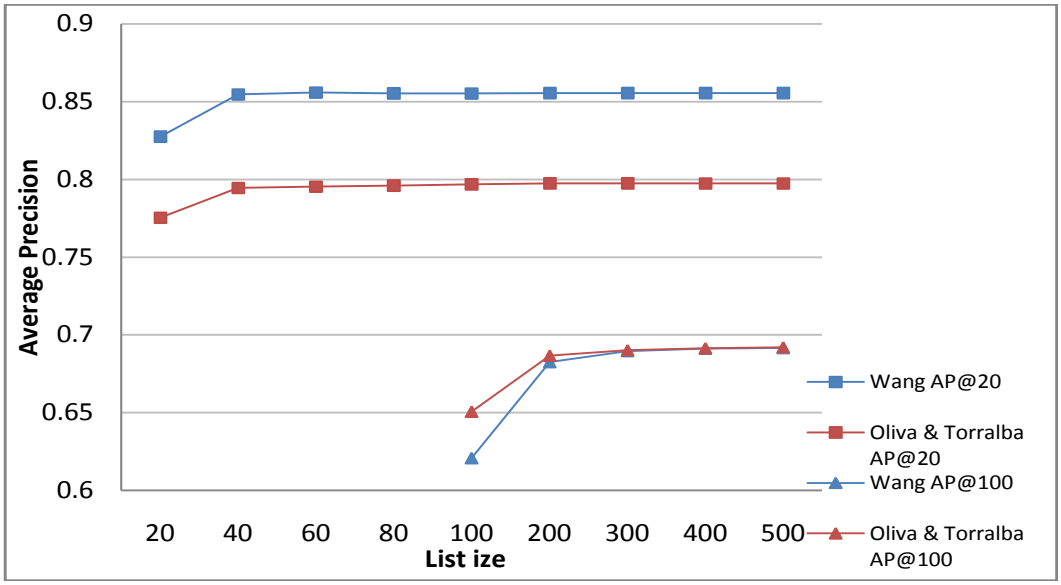


Figure 6.11: Performance variation with the change of the re-rank list size.

Precision at recall was calculated for the Wang and Oliva and Torralba datasets and figure 6.12 gives the precision-recall curves for Wang and Oliva datasets. RB-PRF is quite effective at managing the trade-off between precision and recall, suggesting that searching the initially retrieved list using RB-PRF can lead to a

Table 6.1: Average Precision at 20 (AP@20) with the changing sample size for different evaluation criteria for list size 500.

Evaluation Criterion	05FB	10FB	15FB	20FB
Full ranking	0.8535	0.8547	0.8556	0.8554
Freezing	0.8535	0.8537	0.8445	0.8274
Residual ranking	0.8239	0.8125	0.8016	0.7916

Table 6.2: Average Precision (AP) for different evaluation criteria for a sample size of 15.

Evaluation Criterion	AP@20	AP@40	AP@60	AP@80	AP@100
Full ranking	0.8556	0.8115	0.7782	0.7330	0.6904
Freezing	0.8433	0.8076	0.7765	0.7291	0.6776
Residual ranking	0.7984	0.7595	0.7179	0.6645	0.5920

potential improvement in both measures.

Even though term statistics were not used for first pass retrieval in CBIR-ISIG, we further evaluated the impact of the term statistics on the feedback retrieval performance as it can be used for very large databases. As Log-likelihood was shown to have a high impact on retrieval performance, as demonstrated in Chapter 5, Log-likelihood was used in the feedback search. Figure 6.13 and figure 6.14 show the performance variation with and without the term statistics. Figure 6.13 shows AP@20 with the feedback sample size 10 with varying list sizes. Figure 6.14 shows AP@20 with the feedback sample size 20 with varying list sizes. From these figures it can be seen that there is a slight improvement in retrieval performance. As improvement is fairly small, term statistics were not applied on RF. This will be useful with large data collections.

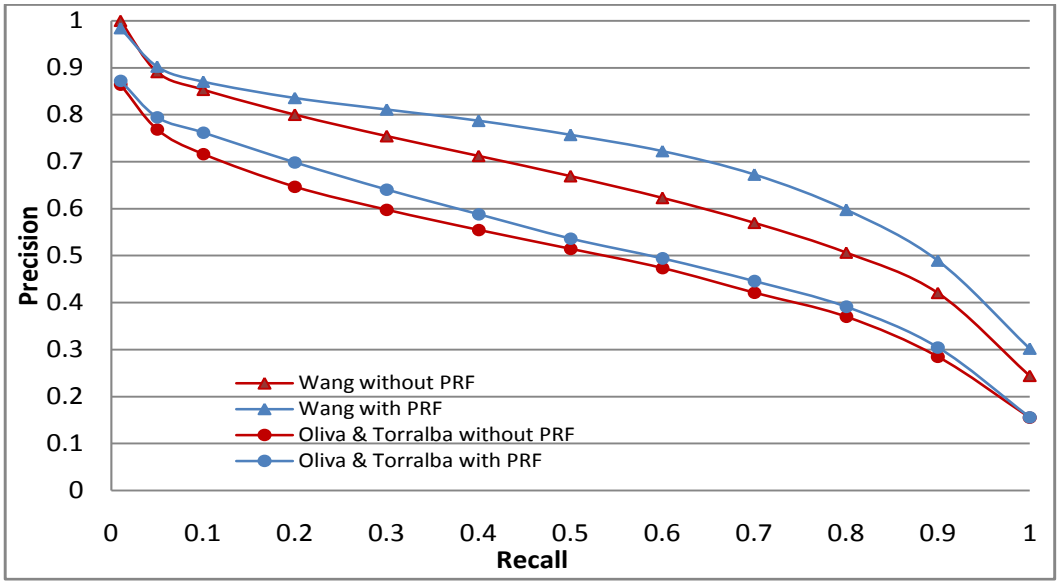


Figure 6.12: Precision recall curve for the Wang and Oliva and Torralba datasets.

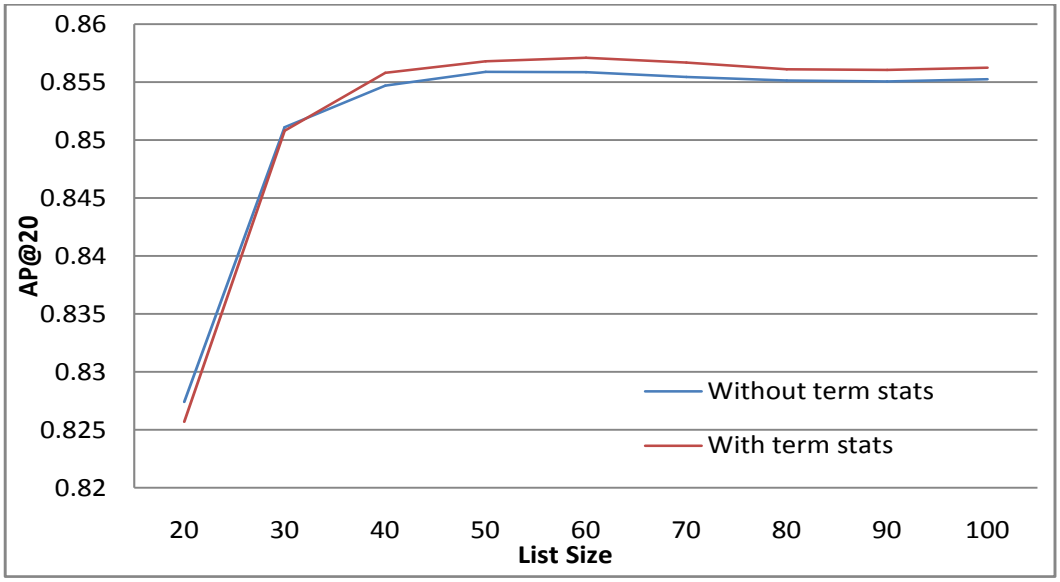


Figure 6.13: Retrieval performance of RB-PRF with feedback sample size 10 on the Wang dataset with and without applying term statistics.

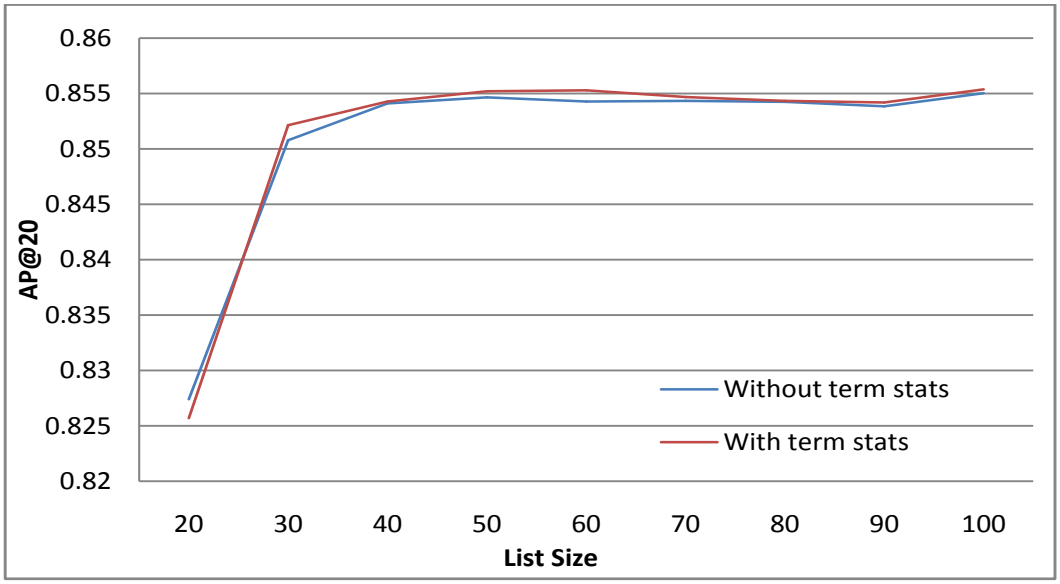


Figure 6.14: Retrieval performance of RB-PRF with feedback sample size 20 on the Wang dataset with and without applying term statistics.

No baseline results were reported in the literature with RF, except [Chowdhury et al., 2012] (2012-in table 6.4 on the Wang dataset).

Tables 6.3, table 6.4 and table 6.5 demonstrate the performance of RB-PRF approach, with baseline CBIR-ISIG system and other baseline systems, which were used to compare the CBIR-ISIG first pass retrieval performance. Table 6.3 shows the results for AP@20 on the Wang dataset and it shows that the RB-PRF approach generates better results using the signature-based approach. Table 6.5 shows the results for AP@50 on the Oliva and Torralba dataset. In these table 6.3, table 6.4, and table 6.5, **bold values** are the highest AP for each class among the compared systems. Underlined values show the highest values CBIR-ISIG system has, among compared systems with no feedback. Furthermore, *italic values* show when RB-PRF is higher with both positive and negative feedback than only with positive feedback. In here we did significance testing using 2-tail t-test with p-value<0.01 (99% significance level) and p-value<0.05 (95% significance level)). In all these three tables ‡ means that the proposed RB-PRF has significant improvement of performance with more than 99% (p-value is less than 0.01) significance level when comparing against our baseline system (CBIR-ISIG). According to ta-

ble 6.3, table 6.4 and table 6.5, it can be observed that the CBIR-ISIG system outperforms the baseline systems on the Wang and Oliva and Torralba datasets and the proposed RB-PRF mechanism works well, with considerable performance improvement.

Figure 6.15 shows the results for some queries with and without feedback on the CBIR-ISIG system.

Except in these datasets, the RB-PRF approach was evaluated on a large dataset (subset of Corel). Table 6.6 shows the results of CBIR-ISIG system with PRF when compared with other systems for AP@20. To compare with the other systems, 500 images were selected at random for evaluation. To remove the effect of biasing, the 15-fold cross validation was used and each time, a different 500 images were selected and the average of the average precision of each round was calculated. Our setup for this experiment was same as the experiments in the other systems [Li et al., 2006, Tao et al., 2007, Bian and Tao, 2010]. Here, images were considered as positive feedback (relevant) if and only if they were in the same class as the query, and others were considered as negative feedback (irrelevant).

In [Li et al., 2006] it was shown that the multi-training SVM method which uses the training technique and the subspace method outperforms several established CBIR RF methods, such as Biased Discriminant Analysis (BDA), Direct kernel BDA (DBDA), and some SVM-based methods. In here, several SVM must be trained. The authors have shown in [Tao et al., 2007] that the proposed kernel-based marginal convex machine by them is better than previously introduced works on CBIR RF, such as SVM, BDA, DBDA, Minimax Probability Machine (MPM), kernel MPM and multi-training SVM. They have considered all the positive examples as one set while the negative examples were split into a number of sets. Each had simple distribution to avoid the problem of excess in the negative examples over the positive examples, but the user always prefer to provide what they want as opposed to what they do not want. Moreover, positive examples were considered as one set and equal importance was given to all the images without considering the importance of each image to the query, which may contribute undue importance to later images in the retrieval list. It was shown that Biased Discriminative Euclidean Embedding is better for RF than other existing methods, which were listed above in [Bian and Tao, 2010]. It modelled both the intra-class geometry and

Table 6.3: Average Precision (AP) of each class along with the whole dataset (Wang) with performance in the literature (AP@20). Here ‡ means that results are statistically significantly greater than 99% significance level when comparing against the baseline CBIR-ISIG system (third column from the last).

Class	2005 [Takala et al., 2005]	2007 [Hiremath and Pujari, 2007a]	2009 [Lin et al., 2009]	2011 [Yuan et al., 2011b]	2011 [Saad et al., 2011]	2013 [Mansoori et al., 2013]	2014 [Hiwale et al., 2015]	2016 [Douik et al., 2016]	CBIR -ISIG	With Positive PRF	Positive & Negative PRF
Africans	0.23	0.48	0.68	0.57	**0.90**	0.68	0.76	0.70	0.75	0.82‡	*0.83*‡
Beach	0.23	0.34	0.54	0.58	0.38	0.28	0.54	0.45	0.68	0.74‡	***0.76***‡
Building	0.23	0.36	0.56	0.43	**0.72**	0.56	0.67	0.59	0.53	0.56‡	*0.60*‡
Bus	0.23	0.61	0.88	0.93	0.49	0.84	0.84	**0.98**	0.86	0.90‡	*0.93*‡
Dinosaur	0.23	0.95	0.99	0.98	**1.00**	0.81	**1.00**	**1.00**	1.00	**1.00**	***1.00***
Elephant	0.23	0.48	0.66	0.58	0.39	0.58	0.70	0.81	0.86	**0.89**‡	0.87
Flower	0.23	0.61	0.89	0.83	0.56	0.55	0.95	**1.00**	0.98	0.99‡	***1.00***‡
Horse	0.23	0.74	0.80	0.68	0.87	0.87	0.94	0.90	0.98	1.00‡	***1.00***‡
Mountain	0.23	0.42	0.52	0.46	0.45	0.48	0.58	0.52	0.77	**0.77**	0.74
Food	0.23	0.50	0.73	0.53	0.87	0.66	0.68	0.61	0.86	0.88‡	***0.91***‡
Average Precision	0.23	0.55	0.73	0.66	0.66	0.63	0.76	0.75	0.83	0.86	***0.87***

Table 6.4: Average Precision (AP) of each class along with the whole dataset (Wang) with performance in the literature (AP@100). Here ‡ means that results are statistically significantly greater than 99% significance level when comparing against the baseline CBIR-ISIG system (third column from the last).

Class	2000 [Li et al., 2000]	2002 [Chen and Wang, 2002]	2008 [Hiremath and Pujari, 2008]	2009 [Banerjee et al., 2009]	2012 [Chowdhury et al., 2012]	CBIR ISIG	With Positive PRF	Positive & Negative PRF
Africans	0.48	0.47	0.48	0.45	0.49	0.52	**0.62**‡	0.60‡
Beach	0.33	0.33	0.34	0.35	0.40	0.46	0.55‡	***0.56***‡
Building	0.33	0.33	0.33	0.35	0.39	0.38	0.39	***0.40***
Bus	0.36	0.60	0.52	0.60	0.58	0.63	0.69‡	***0.74***‡
Dinosaur	0.98	0.95	0.95	0.95	0.96	0.95	0.98‡	***0.99***‡
Elephant	0.40	0.25	0.40	0.60	0.50	0.57	0.67‡	***0.68***‡
Flower	0.40	0.63	0.60	0.65	0.75	0.84	**0.95**‡	***0.95***‡
Horse	0.72	0.63	0.70	0.70	0.80	0.76	0.87‡	***0.89***‡
Mountain	0.34	0.25	0.36	0.40	0.40	0.47	**0.51**‡	0.48‡
Food	0.34	0.49	0.46	0.40	0.51	0.60	**0.67**‡	***0.67***‡
Average Precision	0.47	0.49	0.51	0.55	0.55	**0.62**	**0.69**	***0.70***

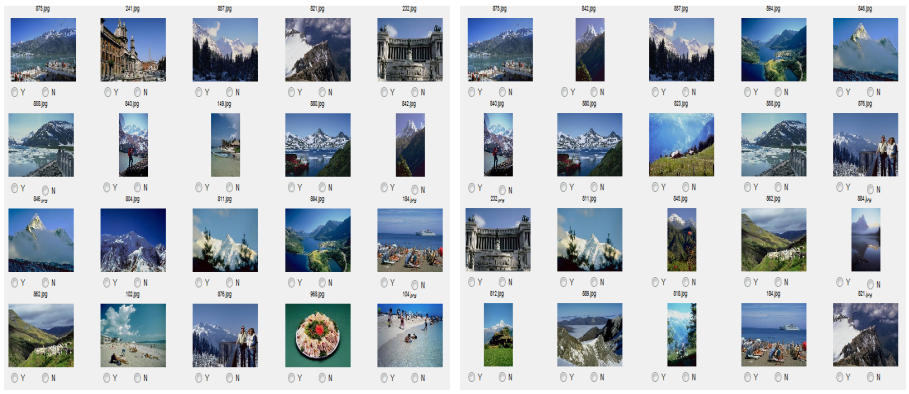
Table 6.5: Average Precision (AP) of each class along with the whole dataset (Oliva and Torralba) with performance in the literature (AP@50). Here ‡ means that results are statistically significantly greater than 99% significance level when comparing against the baseline CBIR-ISIG system (third column from the last).

Class	Region-based [Gokalp and Aksoy, 2007]	CBIR ISIG	With Positive PRF	Positive Negative PRF
Coast (beach)	0.84	0.63	0.72 ‡	***0.79***‡
Country side	**0.50**	0.50	0.48	*0.50*‡
Forest	0.76	0.85	0.96‡	***0.97***‡
Mountain	**0.8**	0.73	0.75‡	0.72‡
Highway	0.62	0.75	**0.81**	***0.81***
Street	0.44	0.94	0.95‡	***0.96***‡
City centre	0.38	0.67	0.76‡	***0.80***‡
Tall buildings		0.73	**0.75**‡	0.73‡
Average Precision	0.62	0.73	**0.77**	***0.79***

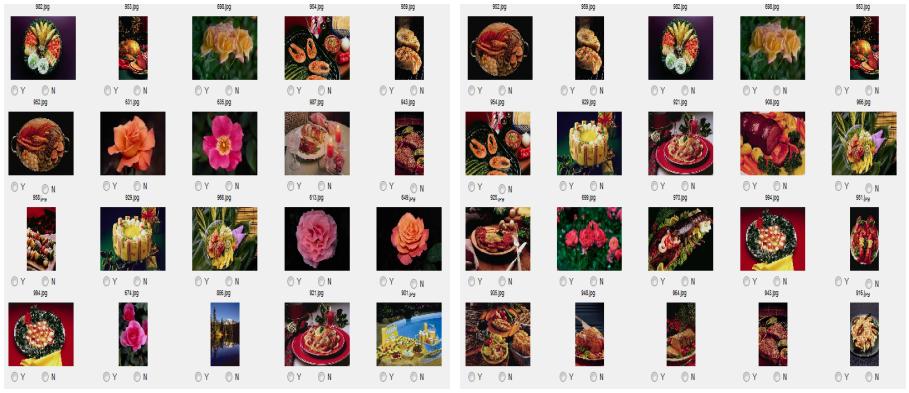
inter-class discrimination. However, there was no control on the relevant scores of the positive samples, as all of them were treated equally. By comparing those methods, we demonstrated that the proposed signature-based PRF outperformed them and rank provides an importance to images according to the similarity which affects the RF.

Table 6.6: Average Precision (AP) of the whole dataset(Corel) with performance in the literature (AP@20). Here ‡ means that results are statistically significantly greater than 99% significance level when comparing against the baseline CBIR-ISIG system

System	No FB	PRF	PRF with Simulated User
MTSVM [Li et al., 2006] (2006)	0.28		0.37
KBMCM [Tao et al., 2007] (2007)	0.28		**0.44**
BDEE [Bian and Tao, 2010] (2010)	0.28		0.37
CBIR-ISIG	**0.29**	0.31	**0.44**‡

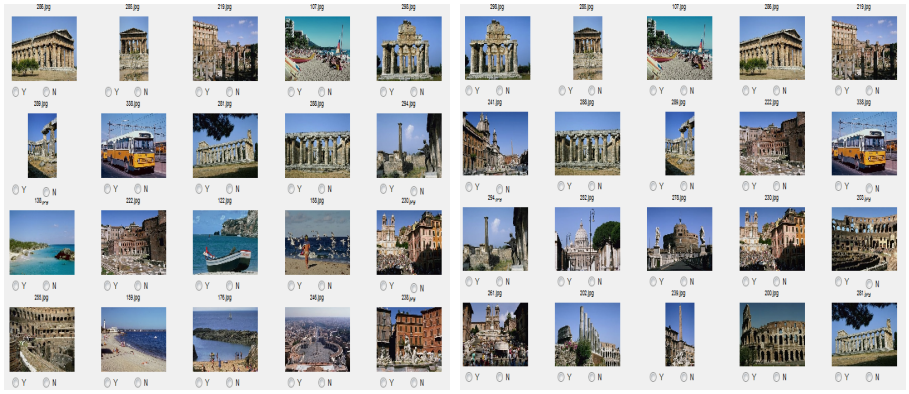


(a) no feedback (13 out of 20) **(b)** with feedback (18 out of 20)



(c) no feedback (13 out of 20) **(d)** with feedback (18 out of 20)

(e) no feedback (19 out of 20) **(f)** with feedback (20 out of 20)



(g) no feedback (12 out of 20) **(h)** with feedback (18 out of 20)

Figure 6.15: The top 20 images for some queries in the Wang and Oliva and Torralba datasets (top left of the no feedback results is the query image).

6.5 Chapter Summary and Conclusions

In this research work a RB-PRF approach is introduced for the application of pseudo RF in CBIR to improve retrieval performance. The original contribution of this research is in the first application of PRF to signatures, taking into account the initial rank order of results to weigh the feedback from each image. This approach has a balance of retrieval speed and quality (see Chapter 8 for the trade-off and speed evaluation). This approach is useful in different situations, such as non-interactive situations and situations where user feedback is not available. The proposed RB-PRF approach performs well and outperforms several systems with previously published results on the same datasets. Our experiments demonstrate that the RB-PRF approach is effective on signature-based representation for image retrieval (system efficiency was evaluated and can be seen in Chapter 8).

This RF was extended for use with interactive RF which will be discussed in Chapter 7.

Chapter 7

CBIR with User Relevance Feedback

Chapter Organisation

The main objective of this chapter is to study how explicit relevance feedback (RF) approach can be used to improve retrieval performance in content-based image retrieval (CBIR-ISIG). Initially, an introduction about RF and a brief overview of the Rank-Based Relevance Feedback (RB-RF) which is proposed in Chapter 6, is presented in Section 7.1 and Section 7.2 respectively. Section 7.3.1 explain the steps to follow in order to apply RF in respect to simulated user feedback (SIM-RF) and real user feedback (URF) accordingly. How the experiments were carried out is presented in Section 7.3.2 and performance and results are analysed in Section 7.3.3. The chapter summary and conclusions are included in Section 7.4. The original contributions discussed in this chapter resulted in publication (vi) in the List of Publications. The combination of the RB-RF approach, proposed in Chapter 6, and the binary image signatures, proposed in Chapter 5, were used with minor modifications to study the impact of real user FB.

7.1 Introduction

CBIR methods allow queries to be specified visually, e.g. through query-by-example methods. In CBIR, image features such as colour, texture, shape, etc, are extracted and indexed automatically with the aim to support the retrieval of images in answer to visual queries. Answering such queries, however, is still a rather challenging task due to the semantic gap that exists between the low-level visual features of images and the semantics that humans associate with objects, entities and scenes in images. Moreover, different users may have different viewpoints about the same set of images. For example, with respect to the images in figure 7.1, a user may feel the two first images are similar because they both represent the Tower Eiffel, while another user may find the first and the third images more similar because in both pictures there a couple posing in front of a monument (or European monument). Furthermore, we can get-rid of the effect of inter-class (figure 7.2) and intra-class (figure 7.3) variability on retrieval performance. In this chapter, we investigate methods for CBIR that aim to address both the semantic gap problem and are flexible to the different viewpoints required by different users.

RF has demonstrated merits in addressing the semantic gap, both for image and text retrieval [Ishikawa et al., 1998, Lu et al., 2000, Yap and Wu, 2007, Wu and Yap, 2006, Yu and Yang, 2001]. In RF, user feedback about the relevancy (or irrelevancy) of documents presented to them is gathered to improve the quality of subsequent rounds of retrieval. This technique has been demonstrated to be effective when applied to CBIR systems also [Yap and Wu, 2007, Wu and Yap, 2006, Zhou and Huang, 2001a, Yu and Yang, 2001, Zhou and Huang, 2001b]. In this chapter, we focus on the use of RF in improving the quality of CBIR systems.

Despite the demonstrations from previous work and the promise of RF applied to CBIR [Yap and Wu, 2007, Wu and Yap, 2006, Zhou and Huang, 2001a, Yu and Yang, 2001, Zhou and Huang, 2001b, Wang et al., 2006, Ko and Byun, 2002], the development of an efficient and effective (both from a user and a system perspective) RF mechanism still remains a challenging and open issue. Pseudo Relevance Feedback (PRF) has been used to remove the burden placed on users by explicit RF mechanisms (see Chapter 6 for more details about these burdens). In PRF, the top-ranked documents obtained from a first pass of retrieval are considered as

Figure 7.1: Examples of different users' viewpoints.

Figure 7.2: Inter-class variability.

implicit relevance indicators. Often also the documents appearing at the bottom of the ranking are considered negative relevance indicators (implicitly irrelevant documents). Most PRF techniques assign the same importance to every document when reformulating the query through the RF mechanism. In Chapter 6, the RB-PRF technique was introduced and we briefly investigated its effectiveness in PRF tasks. In this chapter the technique is briefly re-introduced , but it was extended to explicit RF scenarios and a new host of empirical experiments was performed.

Specifically, in this research work, the RB-PRF technique was expanded and studied in the context of *explicit RF* in CBIR. To this aim, two sets of experiments were instructed : the first was based on simulated interactions, following standard

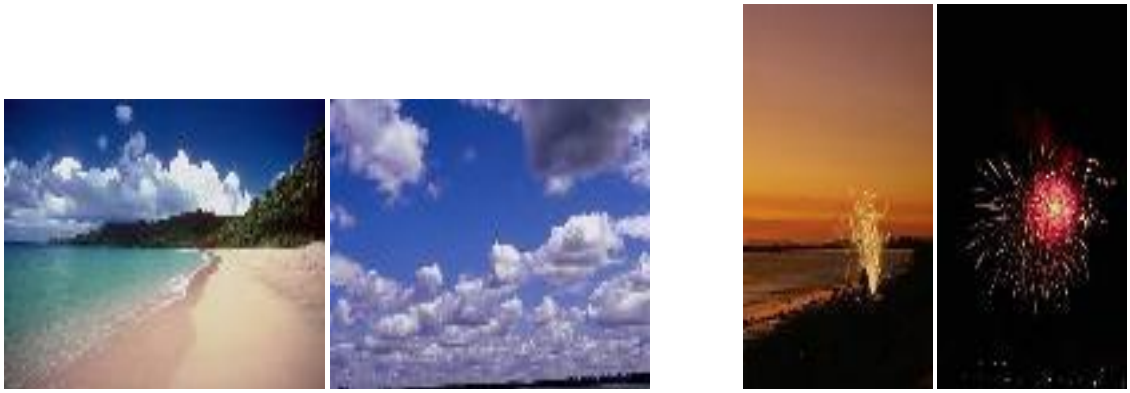


Figure 7.3: Intra-class variability.

practice in the literature; the second was based on a user study where users were asked to interact with our system and provide explicit RF.

The significance of this study is twofold. First, it shows how to use explicit RF effectively with-signature based image retrieval to improve retrieval quality. Second, this approach provides a mechanism for end users to refine their image queries. Unlike text retrieval systems where users are able, and generally prefer, to reformulate their text queries to improve search results, there is no effective way to reformulate an image query. This RF approach provides a solution to this problem. Extensive experiments have been carried out to study the behaviour and optimal parameter settings of this approach. Empirical evaluations based on standard benchmarks (Wang, Oliva and Torralba, and Corel datasets) demonstrate the effectiveness of the proposed approach in improving the performance of CBIR in terms of recall and precision.

7.2 Rank-Based Relevance Feedback

In particular, w controls the granularity of the image similarity and thus provides a means to tackle different users' viewpoints. The larger the value of w, and the more RB-RF promotes images that are visually similar to those selected for RF. In our previous work in Chapter 6 it was found that setting $w = 3$ provided the highest effectiveness in PRF settings, while the preliminary experiments that were performed to investigate RB-RF in the context of PRF experiments in this chapter were carried out in the context of SIM-RF (the experimental methodology

described in Section 6.4). It is found that the highest effectiveness was achieved with higher values of w ($w = 80$ was selected), see figure 7.4. However, these experiments also found that the results are not very sensitive to the value of w which determines how fast the feedback from images decays with rank and this value works well over a wide range of collections and experiments.

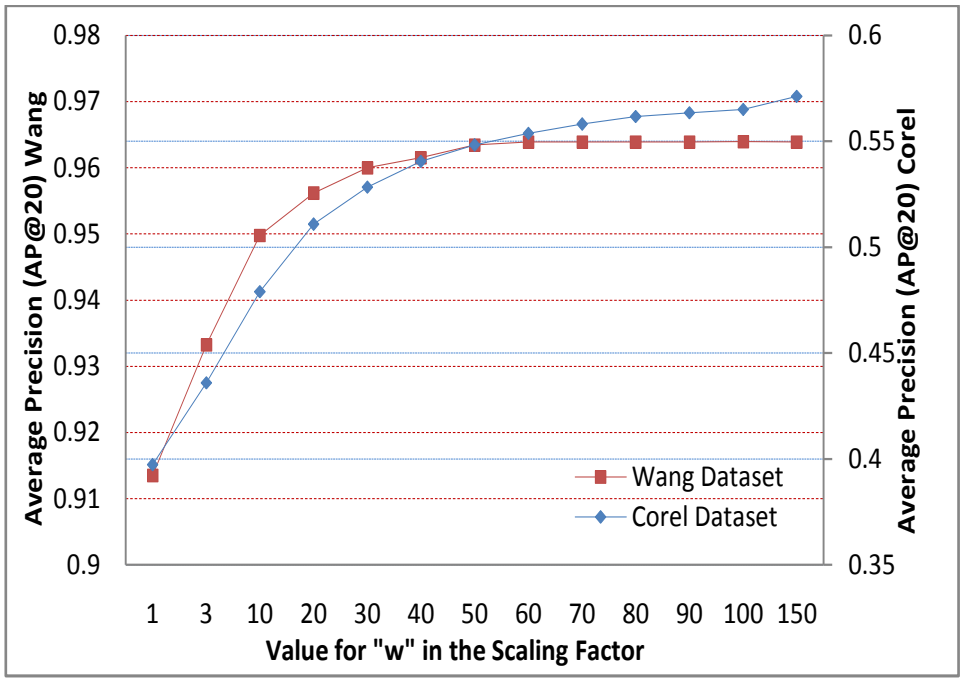


Figure 7.4: Performance variation with the change of w in the scaling factor(MAP@20) on the Wang and Corel datasets for SIM-RF. Here, the feedback sample size is 20 with 8K signature size.

In this research work, we adapted the RB-PRF technique to the settings of explicit RF (RB-RF). Explicit and interactive RF can be used to further address the presence of different users' viewpoints (see figure 7.1), in particular, by narrowing down the image similarity requirements to the characteristics the user is actually after. To this aim, instead of considering the top and bottom N signatures from the list of results retrieved in answer to the original image query, positive and negative feedback vectors were formed using the explicit feedback provided by the user. Apart from the input to the feedback process, the other components and settings of RB-RF remain unchanged in the explicit RF settings. The component of the rank-based (pseudo, simulated and real) RF are summarised in figure 7.5.

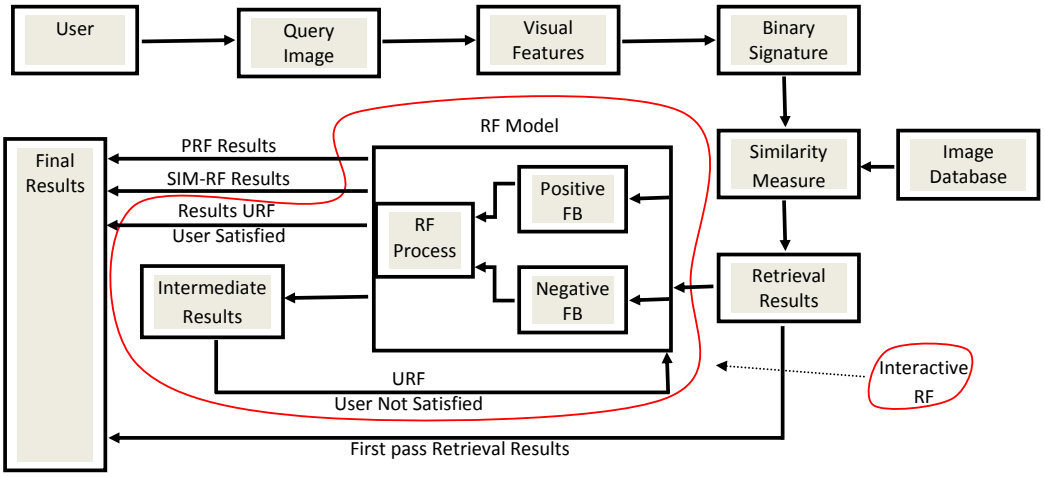


Figure 7.5: Overview of the rank-based (pseudo) relevance feedback system. Lines marked as PRF, SIM-RF (simulated user RF), and After URF (explicit RF) refer to the different methods used to produce the ranked set of retrieved images following RF.

7.3 Evaluation

Next, we aimed to empirically validate our RB-RF mechanism when explicit RF (both positive and negative) is provided by users. To do so, we test our method in two different tasks, a simulated RF and an explicit, interactive RF experiment, using standard datasets and comparing its effectiveness with that of our RB-PRF and other benchmark methods. The details of our empirical evaluation methodology are provided below.

7.3.1 Tasks

Task 1: Simulated RF (SIM-RF)

We first considered the SIM-RF process to evaluate the RB-RF method proposed in Chapter 6. This task was considered because it allows us to compare our method directly with others in the literature [Li et al., 2006, Tao et al., 2006, Tao et al., 2007, Bian and Tao, 2010]. Queries from the considered test datasets (see Section 3.1) were issued to the baseline search system (the TopSig signature-based search system in our case) and a first set of 20 images were formed. RF was simulated by considering as positive feedback the images that belong to the

same semantic class of the query image (see Section 3.1 for details about semantic classes: in brief, this was ground truth information distributed with the image test dataset). Similarly, negative feedback was simulated by considering images that do not belong to the semantic class of the query image. These images were used as a feedback signal to re-rank search results to produce a new set of 20 images (which may be a subset of the 20 initially retrieved images). The re-ranked set of images was then evaluated for relevance using ground truth information distributed with the dataset (the semantic classes). Thus, both the RF and the relevance of the search results are simulated (using collection information) and may not reflect the different viewpoints users may have with respect to their queries. This evaluation task allowed us to compare our baseline CBIR-ISIG system based on the first 20 ranked results with no feedback (the initial results), with systems with PRF and systems with implicit RF (SIM-RF), but no user interaction was considered.

Task 2: Explicit, interacting RF

The second task considered, instead, interaction and feedback explicitly provided by real users. This allowed us to explore the use of RF in real settings and in particular, it allows us to explore whether the proposed technique can adapt to the use of different users' viewpoints. In fact, different users may have different viewpoints about the same image. Sometimes the same user may have different viewpoints about two similar images and may have different viewpoints about the same image at different times. For example, the left image in figure 7.6 shows a beach with clouds in the sky. When using this image as a query, a user may want to retrieve images of skies (e.g. the middle image), while another user may want to retrieve images of beaches (e.g. the right image). The use of RF would help narrow down the user's information need more effectively.

To investigate whether our method addresses the user-viewpoint issue in explicit and interactive RF settings, we extended the CBIR-ISIG user interface to gather user feedback. The user interacted by first issuing an image query to which the system responded with a set of 20 top-ranked images using the baseline search system (CBIR-ISIG). Then the user was asked to select relevant and irrelevant images according to the user's preferences. This feedback signal was issued to the system, which exploits it within our RB-RF mechanism to produce a new set of

Figure 7.6: Visually similar images depending on which viewpoint is selected by the user.

20 images.

In these tasks, 15 users engaged in the system evaluation and they were male and female. No specific group of people were targeted here. Users were from different fields, postgraduate students, computer, electrical, civil and office workers. No specific skill was required, as the system was simple and the user was directed by it. Participation was voluntary. First the task was described to the participants and if they agreed, they were allowed to engage with it.

Note that these experiments were carried out as two steps. In the first step, users were free to select any image from each class as a query. They were given 20 images and the user was asked to select relevant and irrelevant images according to their preference. The system always gave the first 20 images in each iteration. The first 20 images were saturated with relevant images when we ran more iterations. Interaction took place five times (only four RF searches were run) and AP@20 was calculated by using same queries system that was evaluated on SIM-RF (automatic feedback iterations). The result from the first 20 shown images were considered as no feedback results in both cases, as they were the first pass results.

In the second step, the user was given queries and was asked to select relevant and irrelevant images according to preference. In here, when the RB-RF mechanism produces a new set of 20 images, it always showed unseen images in prior iterations. Therefore, no overlap with the set of previously retrieved images was allowed. This was to reflect the real-world interactive settings of this task, where the presentation of the same, redundant results is seen as unlikely to satisfy the user's needs (because the user would have already acquired such images). Interaction took place five times (only four RF searches were run) and only the explicitly relevant information provided by the user on the set of the 100 displayed results

for each query was used to compute the effectiveness metrics. In addition, we also performed a retrieval round using only the baseline system and showed the first 100 results to the user (note that the first 20 results were already shown in the RF process) to gather relevance assessments, so as to allow the comparison of the feedback methods with no feedback baseline.

7.3.2 Evaluation Measures

The RB-RF approach was evaluated on different datasets using different evaluation measures to study the effectiveness of the approach in CBIR by considering retrieval quality. Experiments were performed on several general purpose image datasets Wang, Oliva and Torralba and Corel.

To evaluate the image retrieval approaches investigated in this chapter, we used recall, precision and precision@n (i.e. precision at 20, at 50 and at 100) using equations 3.1, 3.2, 3.3 and 3.4. A Precision-Recall curve is used to demonstrate the system's behaviour with respect to both precision and recall.

Details of these datasets, evaluation parameters and methodologies can be found in Chapter 3.

Note that when considering SIM-RF settings (and PRF), images that belong to the same semantic class of the query image were considered relevant, the others were considered irrelevant. For explicit RF user experiments, relevance information was obtained from the users. We used PRF results here as we needed to compare that with SIM-RF.

To contextualise the effectiveness of the RB-RF method investigated in this chapter, we compared its effectiveness to its PRF version (RB-PRF) (Chapter 6) and the baseline signature-based system (CBIR-ISIG) (Chapter 5) . Furthermore, note that the results of the experiments in Task 1 are directly comparable with methods tested on Wang and Oliva and Torralba datasets.

7.3.3 Results

Precision-Recall curve is generated to compare no feedback, PRF, and SIM-RF. Note that figure 7.7 also shows that the RB-RF method is quite effective at managing the trade-off between precision and recall. Precision and recall improved from no feedback to PRF and to SIM-RF.

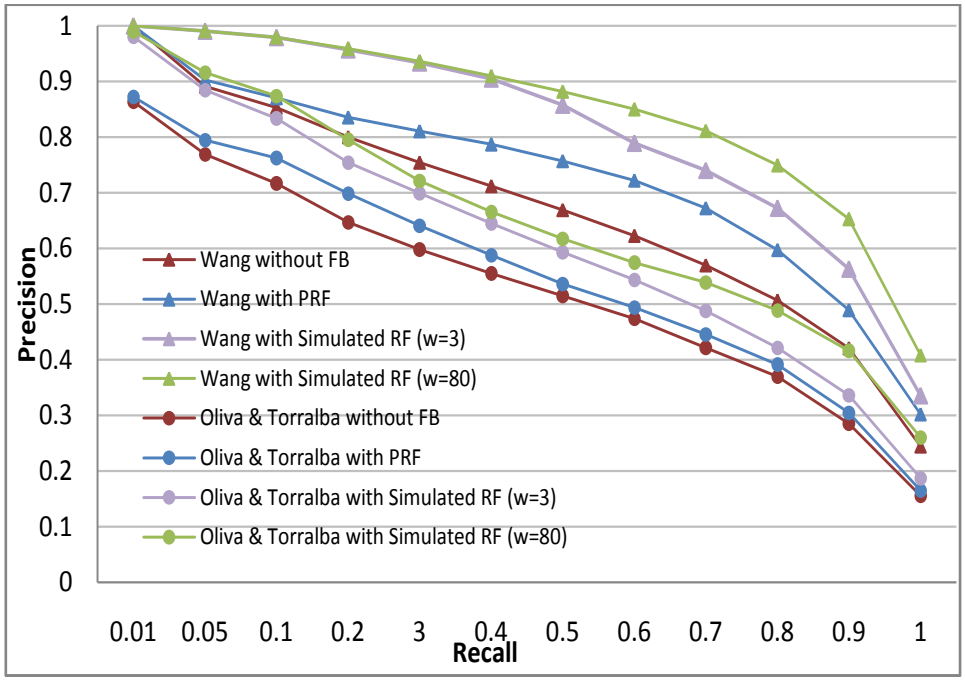


Figure 7.7: Precision-Recall curves for the Wang and Oliva and Torralba datasets.

Table 7.1, table 7.2, table 7.3, table 7.4, table 7.6 and table 7.8 report the retrieval results of the baseline signature-based CBIR (CBIR-ISIG) system and the PRF approach with the simulated user feedback for task 1, while, table 7.6, table 7.7 and table 7.8 report the results using explicit user feedback on task 2. In our earlier work in Chapter 5 we demonstrated that the baseline CBIR-ISIG system outperforms other baseline systems (using AP@20, AP@50, and AP@100 for evaluation).

In table 7.1, table 7.2, table 7.3 and table 7.4, ‡ means that simulated RF has significant improvement of performance with more than 99% (p-value is less than 0.01) significance level when comparing against our baseline system (CBIR-ISIG) and † means more than 95% (p-value is less than 0.01) significance level but less than 99% ($0.01 < \text{p-value} < 0.05$). In table 7.6, table 7.7 and table 7.8 ‡ means that system with RF has significant improvement of performance with more than 99% (p-value is less than 0.01) significance level when comparing against the system without RF.

According to table 7.1, performance of each class of CBIR-ISIG with SIM-

Table 7.1: AP@20 of the Wang dataset. Here ‡ means that results are statistically significantly greater than 99% significance level and † means that that significance level is in between 95% and 99% when comparing against the baseline CBIR-ISIG system

Category	CBIR-ISIG our baseline system (Chapter 5)	PRF RB-PRF (Chapter 6)	SIM-RF (W=3)	SIM-RF (W=80)
Africans	0.75	0.83	$0.89^{\ddagger}$	$0.96^{\ddagger}$
Beach	0.68	0.76	$0.84^{\ddagger}$	$0.90^{\ddagger}$
Buildings	0.53	0.60	$0.70^{\ddagger}$	$0.83^{\ddagger}$
Bus	0.86	0.93	$0.98^{\ddagger}$	$1.00^{\ddagger}$
Dinosaur	1.00	1.00	1.00	$1.00^{\dagger}$
Elephant	0.86	0.87	$0.97^{\ddagger}$	$1.00^{\ddagger}$
Flower	0.98	1.00	$1.00^{\ddagger}$	$1.00^{\ddagger}$
Horse	0.98	1.00	$1.00^{\ddagger}$	$1.00^{\ddagger}$
Mountain	0.77	0.74	$0.93^{\ddagger}$	$0.97^{\ddagger}$
Food	0.86	0.91	$0.95^{\ddagger}$	$0.99^{\ddagger}$
Average Precision	0.83	0.87	0.93	0.97

RF is statistically significantly better compared with the baseline system (CBIR-ISIG) for AP@20 on the Wang dataset except CBIR-ISIG with SIM-RF (w=3) for Dinosaurs class. According to table 7.2, performance of each class of CBIR-ISIG with SIM-RF is statistically significantly better compared with the baseline system (CBIR-ISIG) at AP@100 on the Wang dataset. Table 7.3 demonstrates the results for AP@50 on the Oliva and Torralba dataset and performance of each class of CBIR-ISIG with SIM-RF is statistically significantly better compared with the baseline system (CBIR-ISIG). These table demonstrate how simulated RF work effectively on the CBIR-ISIG system.

Table 7.4 shows the results for AP@20 using a query set size of 500 random images over a 15-fold cross-validation experiment. In this experiment, images are considered as positive feedback (relevant) if and only if they are in the same class as the query image. All other images are considered as negative feedback

Table 7.2: AP@100 of the Wang dataset. Here ‡ means that results are statistically significantly greater than 99% significance level when comparing against the baseline CBIR-ISIG system

Category	CBIR-ISIG our baseline system (Chapter 5)	PRF RB-PRF (Chapter 6)	SIM-RF (W=3)	SIM-RF (W=80)
Africans	0.52	0.60	0.65‡	0.69‡
Beach	0.46	0.56	0.60‡	0.66‡
Buildings	0.38	0.40	0.46‡	0.51‡
Bus	0.63	0.74	0.75‡	0.80‡
Dinosaur	0.95	0.99	0.99‡	0.99‡
Elephant	0.57	0.68	0.75‡	0.83‡
Flower	0.84	0.95	0.96‡	0.98‡
Horse	0.76	0.89	0.86‡	0.89‡
Mountain	0.47	0.48	0.59‡	0.64‡
Food	0.60	0.67	0.74‡	0.78‡
Average Precision	0.62	0.70	0.74	0.78

(irrelevant). Here performance of CBIR-ISIG with SIM-RF (w=80 and w=3) are statistically significantly better with more than 99% significance level compared with the baseline system (CBIR-ISIG) for AP@20 on the Corel dataset. Table 7.5 shows the results of CBIR-ISIG system with SIM-RF when compared with other systems for AP@20. We discussed about this table in Chapter 6 to compare our PRF results. However, as these methods have used simulated user for evaluation we compared with our SIM-RF here and our RF approach demonstrated better performance compared to other systems.

According to figure 7.4, it was found that higher values of w (e.g. w=80) are better for SIM-RF and from preliminary experiments in Chapter 6 in 6.7, lower values (e.g. w=3) are better for PRF. In PRF, we are not aware that those images we considered as relevant and irrelevant are accurately relevant and irrelevant according to the classification. However, the images that we considered as relevant may be irrelevant according to the classification and vice-versa. In

Table 7.3: AP@50 of the Oliva and Torralba dataset. Here ‡ means that results are statistically significantly greater than 99% significance level when comparing against the baseline CBIR-ISIG system

Category	CBIR-ISIG our baseline system (Chapter 5)	PRF RB-PRF (Chapter 6)	SIM-RF (W=3)	SIM-RF (W=80)
Coast (beach)	0.63	0.79	$0.76^{\ddagger}$	$0.81^{\ddagger}$
Country side	0.50	0.50	$0.57^{\ddagger}$	$0.65^{\ddagger}$
Forest	0.85	0.97	$0.97^{\ddagger}$	$0.98^{\ddagger}$
Mountain	0.73	0.72	$0.85^{\ddagger}$	$0.91^{\ddagger}$
Highway	0.75	0.81	$0.89^{\ddagger}$	$0.90^{\ddagger}$
Street	0.94	0.96	$0.97^{\ddagger}$	$0.98^{\ddagger}$
City Centre	0.67	0.80	$0.91^{\ddagger}$	$0.95^{\ddagger}$
Tall Buildings	0.73	0.73	$0.87^{\ddagger}$	$0.97^{\ddagger}$
Average Precision	0.73	0.79	0.85	0.89

Table 7.4: AP@20 of the Corel dataset. Here ‡ means that results are statistically significantly greater than 99% significance level when comparing against the baseline CBIR-ISIG system

Evaluation Criteria	CBIR-ISIG our baseline system (Chapter 5)	PRF RB-PRF (Chapter 6)	SIM-RF (W=3)	SIM-RF (W=80)
CBIR-ISIG	0.29	0.31	$0.44^{\ddagger}$	$0.56^{\ddagger}$

our experiments we assumed the top-ranked images as relevant and the bottom-ranked as irrelevant. Therefore, the scaling factor gives a certain weight to each image which decreases rapidly with the rank and this weight does not have much effect even if selected as mention above. This is the reason that lower values of w show better performance. Therefore, rank-based scaling helps to improve retrieval performance. For the SIM-RF, we considered images relevant if and only if they were in the same class. Therefore, all the images considered relevant were

Table 7.5: Average Precision (AP) of the whole dataset(Corel) with performance in the literature (AP@20). Here ‡ means that results are statistically significantly greater than 99% significance level when comparing against the baseline CBIR-ISIG system

System	No FB	Simulated FB
MTSVM [Li et al., 2006] (2006)	0.28	0.37
KBMCM [Tao et al., 2007] (2007)	0.28	0.44
BDEE [Bian and Tao, 2010] (2010)	0.28	0.37
CBIR-ISIG	**0.29**	**0.56**‡

Table 7.6: User feedback results vs simulated relevance feedback on the Corel dataset -AP@20. Here ‡ means that results with feedback are statistically significantly greater than 99% significance level when comparing against without feedback.

User	No FB AP@20	With FB AP@20	Increment
User RF (w=3)	0.3462 +/- 0.0093	0.6048‡ +/- 0.0186	26%
Simulated RF (w=3)	0.3845 +/- 0.0084	0.6194‡ +/- 0.0087	23%
Simulated RF (w=80)	0.3845 +/- 0.0084	0.6560‡ +/- 0.0096	27%

accurately relevant according to the classification. Then we can give (and here we gave) those images higher weights, which diminishes slowly. Rank affects RF but the impact is less for SIM- RF, as it uses groundtruth. This is the reason higher values of w show better performance in SIM-RF. From these experiments we understood that higher w values are better for SIM-RF (classification) and lower w values are better for PRF. Figure 7.8 shows values of scaling factor (S) with variation of w as a function of i to a get better understanding of the effect of the rank. Here the highlighted lines illustrate how the scaling function varies with the selected w values, 3 (blue line) for PRF and 80 (green line) for simulated and real user FB.

Table 7.7: User feedback results on the Corel dataset. Here ‡ means that results with feedback are statistically significantly greater than 99% significance level when comparing against without feedback.

User	No FB AP@100	User FB AP@100	Increment
User01	0.2864	0.3755‡	9%
User02	0.2806	0.3477‡	7%
User03	0.2678	0.2871‡	2%
User04	0.2528	0.3024‡	5%
User05	0.2878	0.3253‡	4%
User06	0.2712	0.3086‡	4%
User07	0.2861	0.3578‡	7%
User08	0.3011	0.3669‡	7%
User09	0.3048	0.3620‡	6 %
User10	0.2904	0.3535‡	6 %
User11	0.2639	0.3157‡	5 %
User12	0.2766	0.3446‡	7 %
User13	0.2760	0.3650‡	9%
User14	0.2731	0.3213‡	5%
User15	0.3117	0.3584‡	5%
Average Precision	0.2820	0.3395	6%

From that, we concluded that we have to use lesser w values for interactive RF as different users have different viewpoints, and the images a user may judge relevant may not be regarded as being classified into the same semantic class. Moreover, we are not aware of the user's understanding and prior knowledge. Therefore, we used $w = 3$ for interactive RF.

To evaluate the image retrieval accuracy of the proposed RF approach, the system was evaluated using real users. Table 7.6, table 7.7 and table 7.8 show the experimental results with real user feedback. A total of 15 users were recruited. Each user was asked to submit 83 queries and interact with the search system during five iterations. These experiments were carried out in two steps as described, and table 7.6 shows the results from the first step, while table 7.7 and table 7.8

Table 7.8: User feedback results vs simulated relevance feedback on the Corel dataset -AP@100. Here ‡ means that results with feedback are statistically significantly greater than 99% significance level when comparing against without feedback.

User	No FB AP@100	With FB AP@100	Increment
User RF (w=3)	0.2820 +/- 0.0161	$0.3395^{\ddagger}$ +/- 0.0301	6%
Simulated RF (w=3)	0.2657	$0.2879^{\ddagger}$	2%
Simulated RF (w=80)	0.2657	$0.2934^{\ddagger}$	3%

shows the results from the second step.

Table 7.6 compares the AP@20 without feedback and that was obtained after the fifth iteration. According to table 7.6, the results are statistically significantly better. Improvement is higher with real user feedback than SIM-RF for $w = 3$ and 1% higher for $w = 80$. However, these results cannot be directly compared, as one is based on user perception while the other on classification.

Next, we compared the results obtained across the rounds of RF interactions with those obtained when users were asked to assess 100 search results initially retrieved by the baseline system (without RF). For this we compared approaches using AP@100. This evaluation is quite similar to residual ranking. We removed the top-ranked signatures from the ranked set after they were used to train the RF system and they were added to the final ranking as it was evaluated. Table 7.7 shows the results from each of the 15 users recruited in the experiment, while table 7.8 compares the AP@100 without feedback with that obtained after the fifth feedback iteration for both the explicit interactive method and the simulation. Even though RB-RF delivers only a 6% improvement in search effectiveness when users feedback is used, this is statistically significantly better than no feedback. In addition, the improvement is higher for interactive user RF than for SIM-RF. From these experiments, it can be concluded that the user's viewpoint is different from the classifications encoded in the semantic classes of datasets like Corel, Wang,

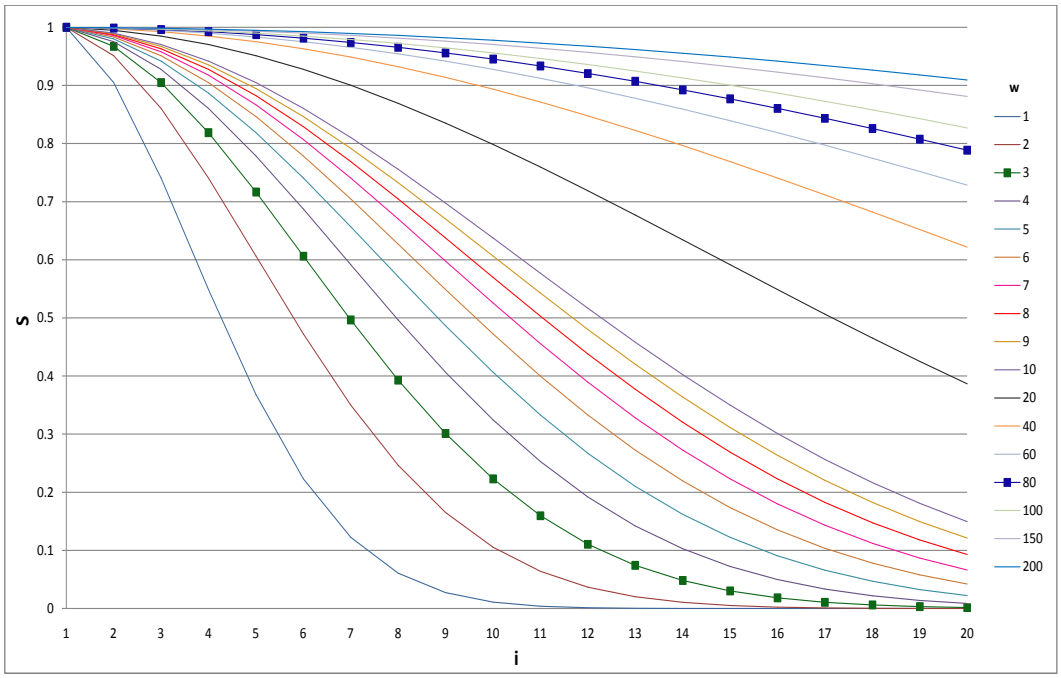


Figure 7.8: Scaling factor S variation with w as a function of i.

and Oliva and Torralba, and user interaction is essential to obtain better retrieval results.

7.4 Chapter Summary and Conclusions

This research work examined an RB-RF approach and extended that work to the situations in which SIM-RF and explicit, interactive user feedback were provided. This mechanism help users to refine their query image when use is unable to reformulate an image query. Initially, parameters were studied for the SIM-RF context, then the settings and parameters that were studied for the PRF were used to evaluate the resulting CBIR system with RB-RF in a set of interactive user experiments. The empirical validation of the proposed method was carried out on three benchmark datasets and suggested that the user behaviour measured in the real, explicit and interactive user experiments is different from that of simulated users, in particular with respect to the differing users' viewpoint exhibited by real users. With this regard, our method has exhibited scalability and effectiveness to the change of users' viewpoints. Moreover, both the SIM-RF and interactive RF

results are statistically significantly better (at 95% significance level) than first pass retrieval results in the CBIR-ISIG system.

Chapter 8

Scalability of the Content-Based Image Retrieval System

Chapter Organisation

The main objective of this chapter is to evaluate the scalability of the CBIR-ISIG system and the RB-RF mechanism. Thus, scalability of the binary signature-based approach in Chapter 5, RB-RF approaches in Chapter 6 and Chapter 7 in CBIR were investigated here. Initially, in Section 8.2, the effectiveness of the system is evaluated, followed by efficacy of the system in Section 8.3. Then the robustness of the system is evaluated and explained in Section 8.4. The chapter summary and conclusions are included in Section 8.5.

8.1 Introduction

In CBIR, scalability is quite a challenging issue. We considered scalability in the sense of effectiveness, efficiency and robustness. Initially, a CBIR system must be able to conduct semantic image retrieval and that can be defined, as effectiveness of the system. In here, we discuss the effectiveness of our CBIR-ISIG system on different standard datasets and furthermore, the effect of RB-RF is evaluated to show its image retrieval ability.

The next factor we will discuss here is the system's efficiency. This is very important for interactive retrieval. First pass retrieval time and feedback search

time were evaluated using different database sizes and feedback sample sizes. We then examined the pre-processing time of a query image as it is essential when the query image from the outside. From these experiments, we show the efficacy of the system by considering the pros and cons.

Then we will discuss the robustness by considering different alterations. Here. we consider different alterations in different ranges. Figures signify the system's robustness to various image degradations.

The main observations from the experiments and the conclusion are discussed in the conclusion section.

8.2 Effectiveness

Effectiveness measures the accuracy (quality of search) of a system. That means how effectively the system achieves the correct results. The effectiveness of a CBIR system can be measured by two methods, either based on classification data or from the user's perspective. In classification, images are considered relevant if and only if they are in the same class, and from user's perspective, images are considered relevant if a user says they are relevant. The CBIR-ISIG system achieved significant

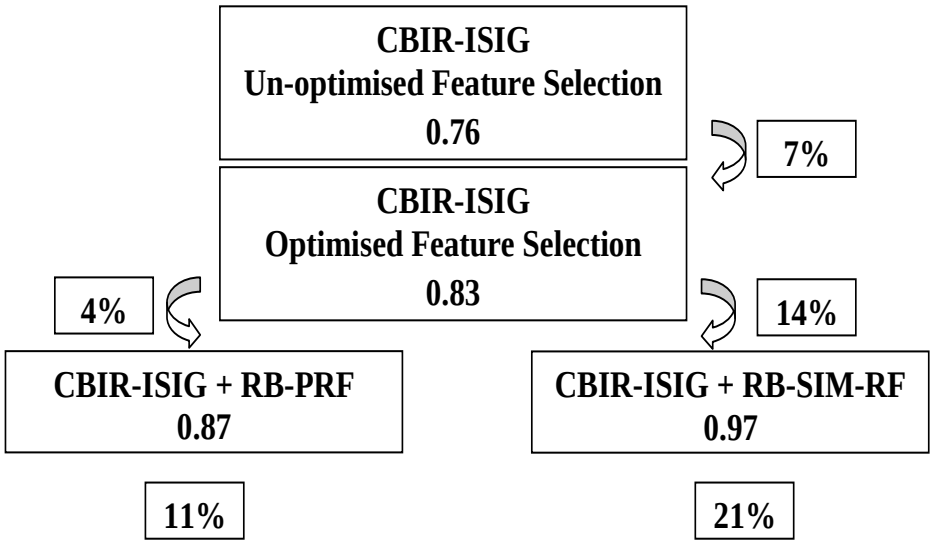


Figure 8.1: Improvement of retrieval performance over different steps in CBIR-ISIG - AP@20 on the Wang dataset.

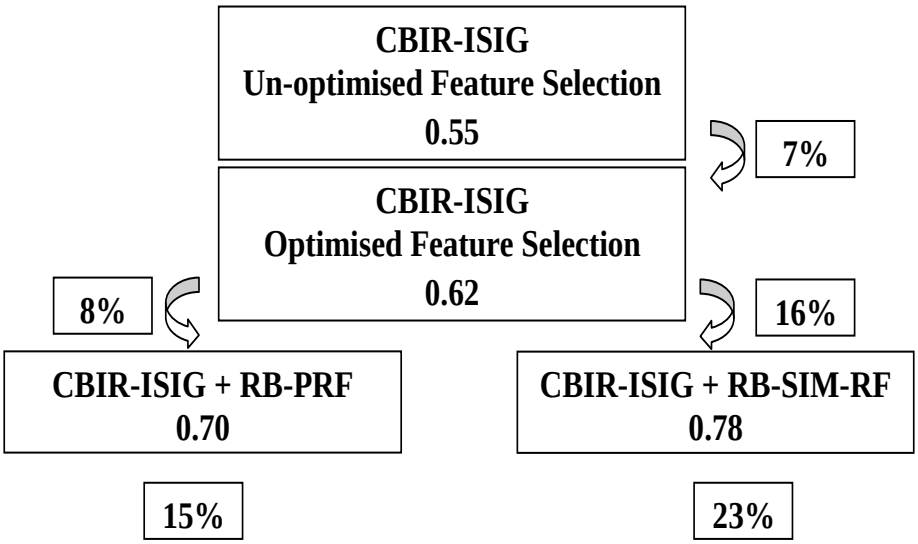


Figure 8.2: Improvement of retrieval performance over different steps in CBIR-ISIG - AP@100 on the Wang dataset.

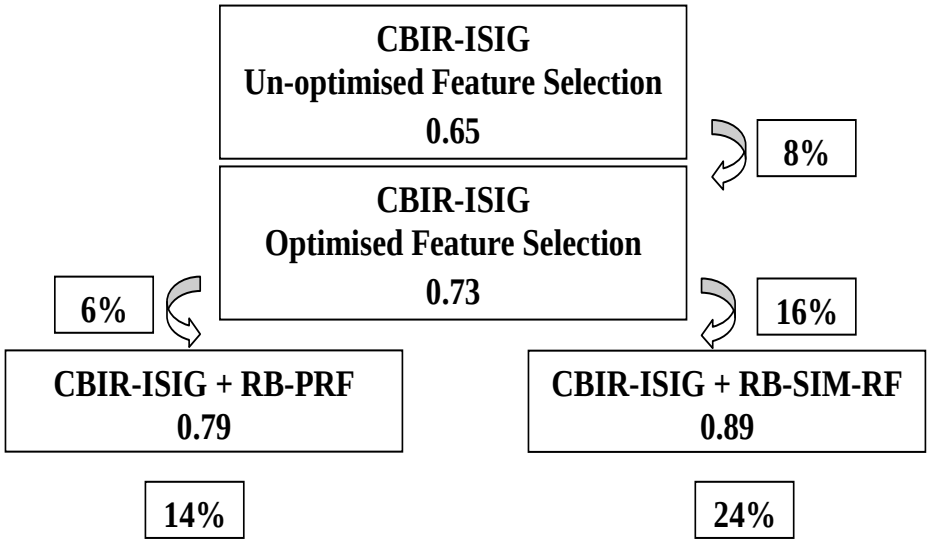


Figure 8.3: Improvement of retrieval performance over different steps in CBIR-ISIG - AP@50 on the Oliva dataset.

performance in retrieval accuracy and the system was empirically evaluated using several datasets. A brief overview of the effectiveness of the CBIR-ISIG system can be seen from the figure 8.1, figure 8.2, figure 8.3, figure 8.4 and figure 8.5.

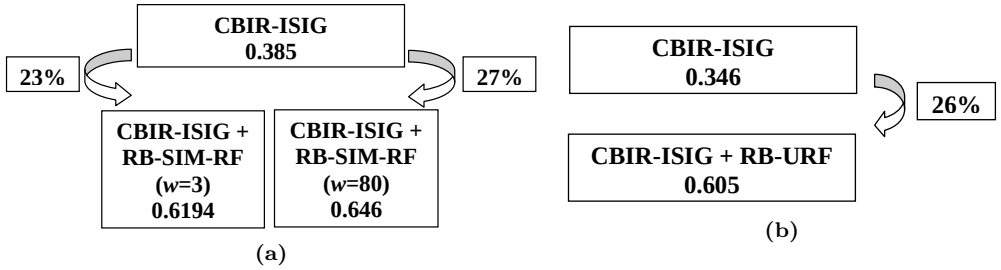


Figure 8.4: Improvement of the CBIR-ISIG for AP@20 with the feedback (a)Simulated User (SIM-RF) and (b)Real User (URF) on the Corel dataset.

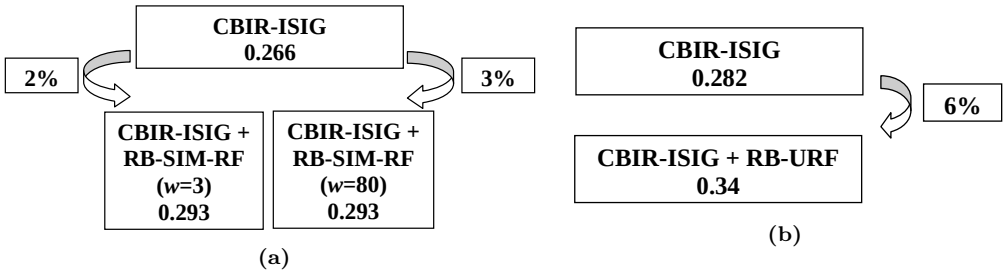


Figure 8.5: Improvement of the CBIR-ISIG for AP@100 with the feedback (a)Simulated User (SIM-RF) and (b)Real User (URF) on the Corel dataset.

Figure 8.1 and 8.2 shows the retrieval accuracy on the Wang dataset, AP@20 and AP@100 respectively. Overall, CBIR-ISIG with PRF achieves an 11% improvement over the baseline system, while it is 21% with RB-SIM-RF for AP@20. Moreover, CBIR-ISIG with PRF achieves a 15% improvement over the baseline system, while it is 23% for RB-SIM-RF for AP@100. Precision at each step is indicated in the boxes.

Figure 8.3 shows the retrieval accuracy on the Oliva and Torralba dataset for AP@50. Overall, CBIR-ISIG with PRF achieves 14% improvement over the baseline system, while it is 24% with RB-SIM-RF.

Figure 8.4 and 8.5 shows retrieval accuracy on the Corel dataset. The results for SURF and URF shown here are the results achieved after 5 iterations. Precision at each step is shown in the boxes. According to figure 8.4, improvement is nearly similar in both cases, while figure 8.5 shows that URF achieves twice the improvement with SURF for AP@100. The reasons to use different values for w

are described in Chapter 6 and 7. However, the CBIR-ISIG system shows better retrieval performance on the Corel dataset.

The effectiveness of the system has been empirically evaluated in each section and the evaluation methodology and results have been described in the Evaluation section in each chapter. From the evaluation results. it can be concluded that the CBIR-ISIG system has a good retrieval performance and the RB-RF approach improves it further.

8.3 Efficiency

A CBIR system must be efficient in order to achieve results without letting user feel the elapse time (without taking too much time). Traditionally, CBIR systems are characterised by a slow response time or high computational costs due to the high dimensionality of the feature spaces used to describe images. This problem has been tackled by a number of CBIR techniques for dimensionality reduction [Yang et al., 2015b,Torralba et al., 2008b,Feng et al., 2016,Lv and Wang, 2013,Shu et al., 2015,Gallas et al., 2015,Liu et al., 2014] and the CBIR-ISIG system used RI, which is an incremental approach that has a low computational cost, less complexity and better accuracy with the combination of the TOPSIG system [Geva and De Vries, 2011]. The use of signature-based methods, which are designed to make efficient retrieval of high-dimensionality data feasible, allows our system to work quickly and achieve retrieval times in the milliseconds through the use of scalable signature search approaches [Chappell et al., 2015].

The RF process can also be performed efficiently due to the fact that the signatures for the retrieved images remain resident in memory after the initial search and performing the re-ranking only requires the calculation of a small number of Hamming distances and the sorting of the results according to the updated distances. Therefore, the time to perform the final PRF step is negligible (in ms scale).

The run-time efficiency of the retrieval methods was examined. First, the CBIR-ISIG first pass retrieval speed was computed. Figure 8.6 reports the retrieval run-time in milliseconds (ms) for the full Wang, Oliva and Torralba, and Corel datasets for different signature sizes. The average time was calculated by running the search several times so as to average the noise, due to other overhead

processes at run-time. However, the effect of the overhead may have a lot more impact on smaller datasets than they would on a large dataset and that can be evidenced by comparing figure 8.6 and figure 8.7. Figure 8.7 reports the retrieval run-time for increasingly larger samples of the Corel dataset. Even though the time required for the first pass of retrieval time increases with the increase of the image database size, the trend is logarithmic.

All these searching times were computed only on a searching phase. the system was able to achieve this speed when the query image was taken from the database as the pre-processing step was not required (as the query signature is already in the signature database). However, pre-processing was required when the query image was taken from outside from the database. In pre-processing, feature extraction, the assignment of cluster centres and the generation of BoW representation were encountered. Signature generation was considered at the time of searching (figure 8.6 and figure 8.7). The time for the assignment of cluster centres(0.0094s) and the generation of BoW representation (0.0131s) was negligible compared with the time of feature extraction (1.3534s) per image. In our computational configuration, the average time which was required from feature extraction (from 256 by 256 image) to the generation of symbolic representation was 1.3759s. In here the average time was calculated by running the search several times so as to average the noise. However, pre-processing time was significantly higher than the searching time. Therefore, searching time will be in seconds if the query image is new as the retrieval time is the pre-processing time + searching time for image searching. The pre-processing time is only dependent on the query image size, as the other feature settings are not changed after generating the CBIR system, and the searching time is dependent on the database size and the signature size. Image down sampling is an answer to the large image sizes and this may lead to deterioration of the performance.

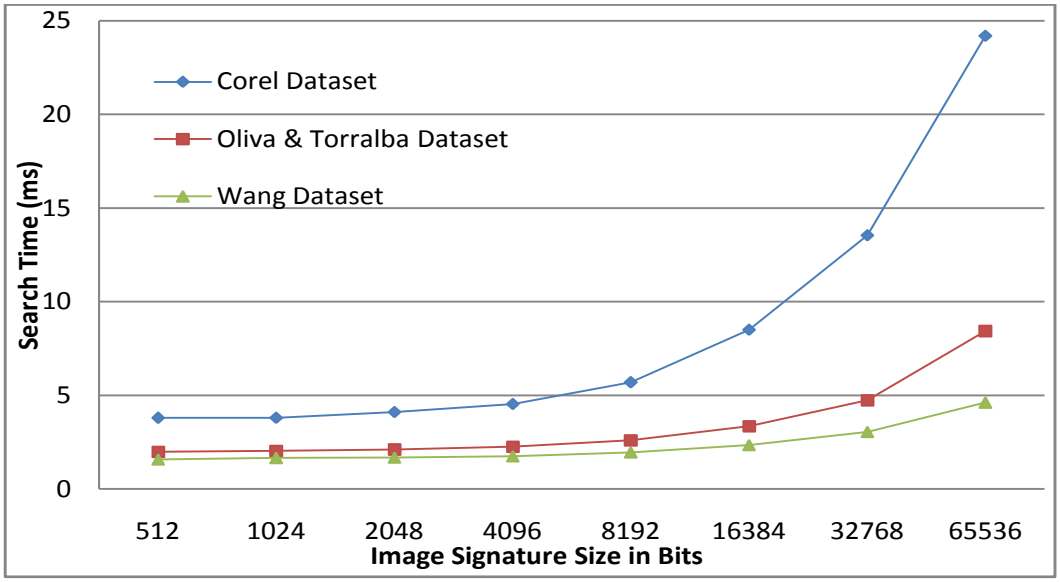


Figure 8.6: Search time vs signature size on the Wang, Oliva and Torralba and Corel datasets.

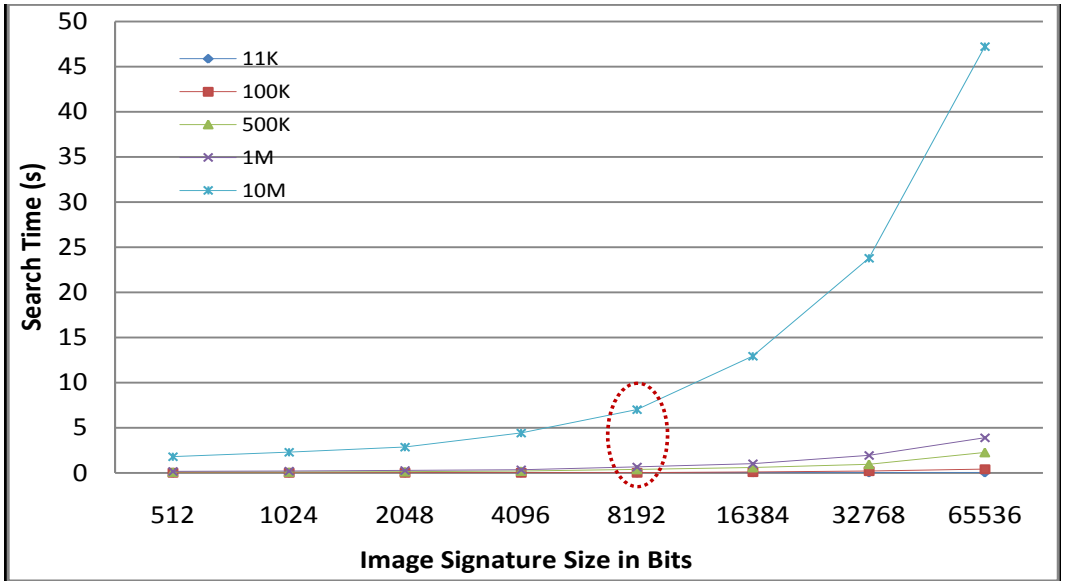


Figure 8.7: Search time (first pass retrieval) vs database size.

As we wanted to study the trade-off between effectiveness and efficiency, search time and average precision were calculated over a range of signature sizes. Figure 8.8, figure 8.9 and figure 8.10 show evaluation results on Wang, Oliva and

Torralba, and Corel datasets respectively. From these figures it can be seen that retrieval time rapidly increases after an 8192 signature size but not the retrieval quality. Instead the results deteriorated slightly. Therefore, it can be concluded that our choice of an 8192 signature size is the most suitable size for the CBIR task detailed here. However, the signature size must be selected by considering which factor is the most important for the system in the correlation between speed and quality.

Figure 8.11 reports the RF retrieval run-time, which includes the time needed to generate the feedback signatures and the time taken to search and re-rank the result list. In conclusion, it can be stated that the retrieval time increases when the signature size and feedback sample size increase. It must be noted that the RF retrieval run-time does not depend on the dataset size ,as the retrieval results are resident in memory; this time is negligible when compared with the retrieval time for large datasets. Thus, the run-time required by the RF process scales to increasing datasets and, indeed, it is negligible, regardless of the collection size.

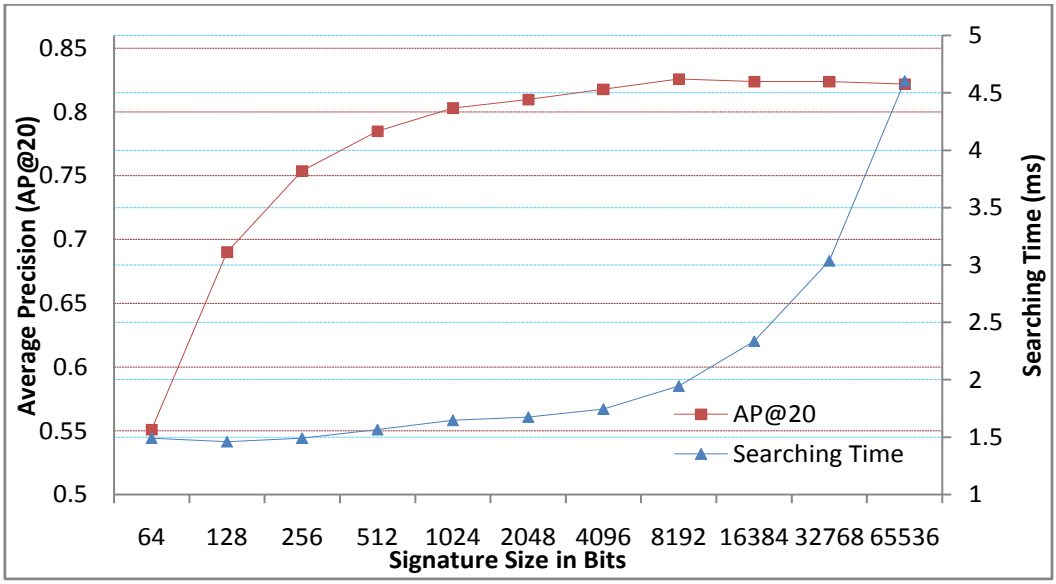


Figure 8.8: The trade-off between retrieval quality and speed on the Wang dataset.

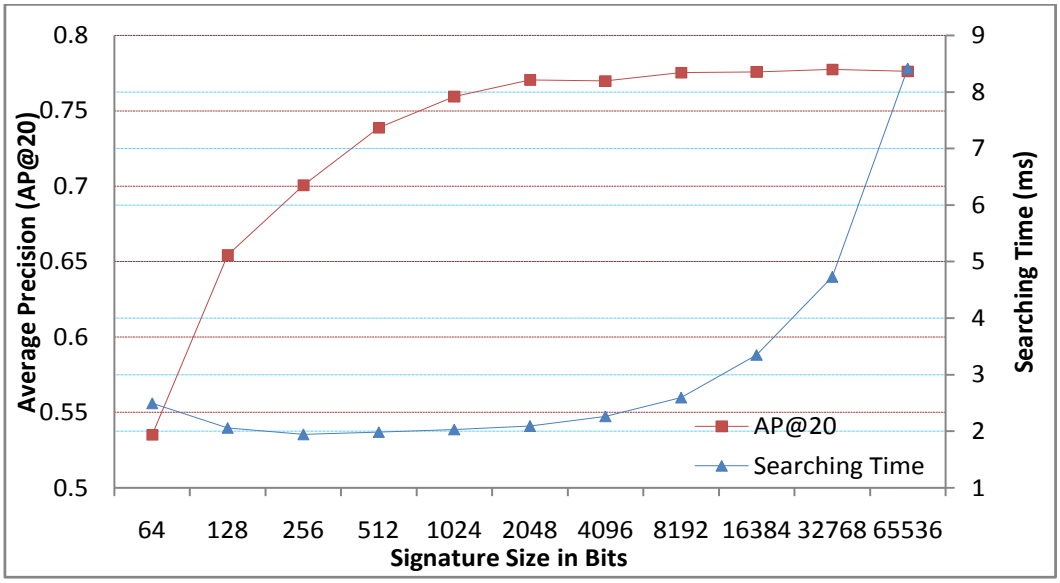


Figure 8.9: The trade-off between retrieval quality and speed on the Oliva and Torralba dataset.

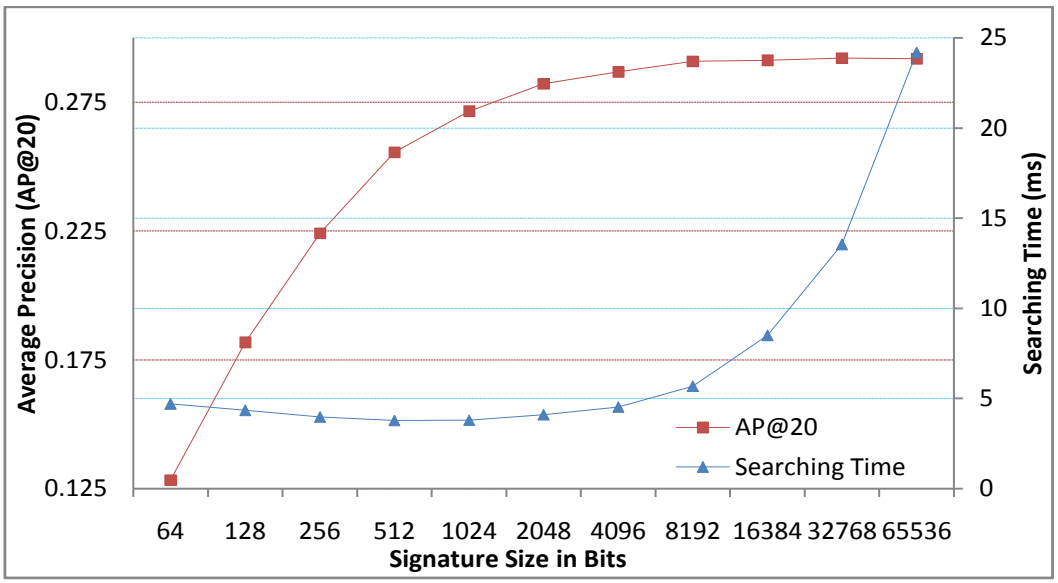


Figure 8.10: The trade-off between retrieval quality and speed on the Corel dataset.

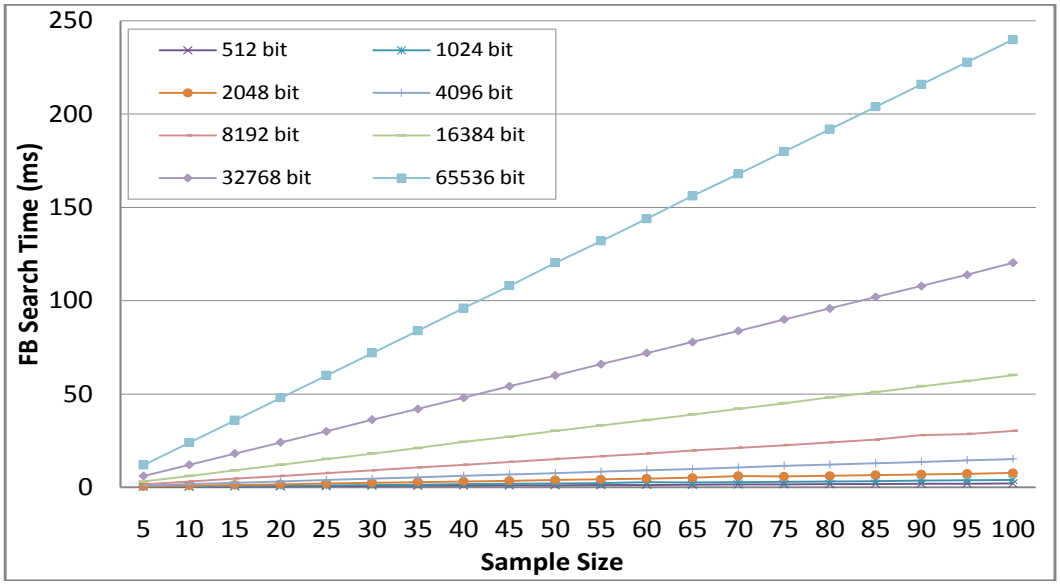


Figure 8.11: Search time vs sample and signature size (feedback signature generation and search time for a 500 list size).

Sample Size		PRF 500 list			SIM-RF 500 list	
	Wang	Oliva	Corel	Wang	Oliva	Corel
15FB	0.864	0.815	0.317	0.924	0.902	0.370
20FB	0.865	0.814	0.317	0.929	0.909	0.381
50FB	0.864	0.818	0.316	0.933	0.918	0.416
100FB	0.859	0.816	0.315	0.929	0.919	0.444

Sample Size		PRF Full list			SIM-RF Full list	
	Wang	Oliva	Corel	Wang	Oliva	Corel
15FB	0.851	0.795	0.305	0.910	0.859	0.369
20FB	0.844	0.788	0.304	0.914	0.863	0.380
50FB	0.841	0.786	0.300	0.911	0.860	0.415
100FB	0.834	0.780	0.295	0.902	0.852	0.443

Table 8.1: AP@20 on the Wang, Oliva and Torralba, and Corel datasets for changing re-rank list size (smaller re-rank list size vs full list).

Moreover, it must be noted that re-searching the entire collection is an unnecessary computational cost and doing so leads to worse overall results, as extra details may add non-relevant results in the searching process. Table 8.1 provides the results for AP@20 on the Wang, Oliva and Torralba, and Corel datasets. It shows the results for a feedback re-rank list size of 500 and a full database over different feedback sample sizes. It can be seen that a smaller re-rank list size provides better results, as mentioned. Therefore, this proposed feedback searching process is efficient and effective.

It must be noted that these searching times in the thesis were computed using a system with Intel(R) Xeon (R) E5-2665s clocked at 2.40 GHz. The run times recorded in our experiments could be contextualised and contrasted with results reported in prior work. However, a direct comparison is not possible because the reported results were obtained using a different hardware architecture and experimental settings and the implementations of previously proposed methods are not freely available. Table 8.2 and table 8.3 compare search time with existing systems while table 8.2 shows first pass retrieval time and table 8.3 shows feedback search time. It must me noted that these first pass searching times (table 8.2) of all the systems were computed on Linux OS. Normally, all these systems used 20 images as feedback in table 8.3.

When compared with other methods [Chen and Wang, 2002, Carson et al., 2002, Li et al., 2000] in table 8.2, our system is faster. Although this might be due to the fact that the experiments detailed in other papers had used limited hardware when compared with the one we used in our experiments but they also had used smaller datasets for comparison. Therefore, this limitation is traded-off by the small dataset sizes. The hardware used by [Torralba et al., 2008b] is closer to our hardware settings because it is multi-threaded on a quad core processor. While again the results are not directly comparable because of the other settings used, we can note that the collection size and signature size is directly comparable.

In table 8.3, although the results are not directly comparable, the architecture of [Su et al., 2011b] seems to be similar. Thus, our system seems to be faster. However, system [Qian et al., 2016] has not provided the system specification.

Although we cannot have an ultimate conclusion, we can speculate that if the systems are compared within the same architecture, our system will demonstrate

System	Model	Specification	Database Size	Search Time(s)
[Chen and Wang, 2002]	Unified Feature Matching	Pentium III 700MHz	60 000	0.7
[Carson et al., 2002]	Blobworld	Pentium III 700MHz	35 000	5.3
[Li et al., 2000]	Integrated Region Matching	Pentium Pro 430MHz	200 000	1.5
[Torralba et al., 2008b]	Restricted Boltzmann Machines	Quad-core processor		
	256 bit	Multi-thread	12 900 000	0.23
Our method	Signature based	Intel Xeon Quad-core processor		
	256 bit	Multi-thread	12 900 000	0.031
	2512bit	Single-thread	1 000 000	0.176
	8K bit	Single-thread	1 000 000	0.676

Table 8.2: First-pass search time comparison with other systems.

System	Model	Specification	Database Size	Search Time(s)
[Su et al., 2011b]	Navigation Pattern based RF	Intel Dual Core Xeon 2.13 GHz	6 000	0.897
			8 000	1.178
			10 000	1.479
[Qian et al., 2016]	QPM-Cluster based		101 240	1.280
Our method	Rank-based RF	Intel quad core Xeon 2.40 GHz	11 000	
	8K bits			0.006
	65K bits			0.047

Table 8.3: Feedback search time comparison with other systems.

faster retrieval results. In addition, we note that the searching time of our RB-RF method is not affected by increasing the size of the database.

8.4 Robustness

The proposed signature-based image retrieval system used a range of image descriptors covering all the features and it was based on local representation. Therefore, we expected the system to be robust to some extent. Then the CBIR-ISIG system was evaluated to illustrate the robustness by taking a random ten images from each class from the Wang dataset. The system was tested for several image alterations, namely horizontal and vertical flipping, horizontal and vertical shifting, cropping, rotating, sharpening, blurring, saturation variations, brightening, darkening and shape distortion by adding noise. The goal was to demonstrate the ability of the system to retrieve the unmodified image when its altered version is given as the query image. Initially, all these selected images were modified with the above mentioned changes and new database was made. Then a retrieval list was taken for each image in the generated database by providing each image as a query image. Then the rank of the unmodified modified image was considered and an average was calculated for each case.

Figure 8.12, figure 8.13, figure 8.14, figure 8.15, figure 8.16, figure 8.17, figure 8.18, figure 8.19, figure 8.20, figure 8.21, figure 8.22, figure 8.23 and figure 8.24 show the change of average rank over a range of degradations. Each alteration was studied under a range of variations.

We defined system robustness as follows: the system is robust if it is able to retrieve the original image within the first five images of the retrieved list when the modified version of that image was provided as the query image. If so, our system was extremely robust to flipping and the unmodified image always ranked first; it is not shown in these images. The system was also robust to horizontal shift, darkening, sharpening, high saturation and pixel change on the top, bottom and diagonal. Moreover, system was robust to vertical shift up to 30 pixels on both sides, horizontal and vertical shift up to 20 pixels and again at 125-150. Furthermore, the system was robust to 30% brightening, 40% less saturation, 25% cropping, 10° rotation, blurring with a filter size of 15, 4900 random pixels and a 90*90 pixel change in the middle. The system may be more robust to rotation

than shown here, as these results were taken by manual rotation and four black triangles were introduced when rotating. This might have affected the performance inversely.

Examples for some cases are shown in figure 8.25. It shows some query examples with the first five matches for different alterations of the images.

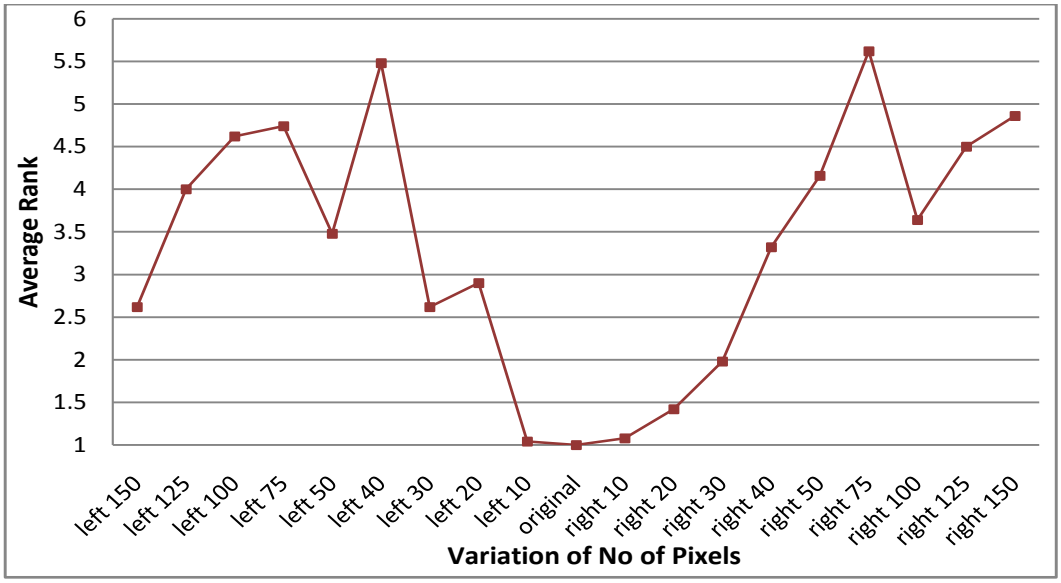


Figure 8.12: Average rank vs image horizontal shift. Here the pixels were shifted right and left.

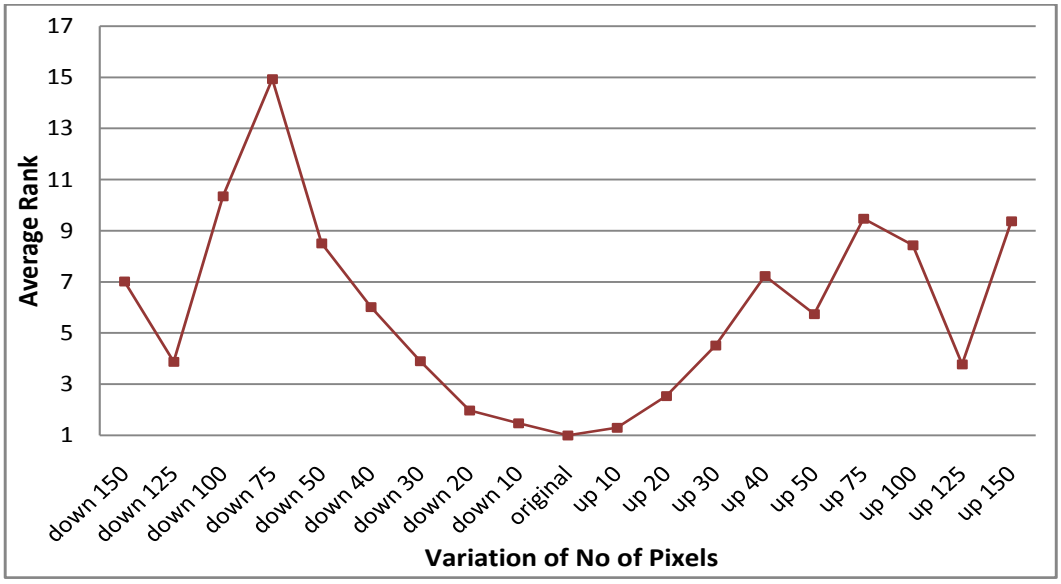


Figure 8.13: Average rank vs image vertical shift. Here the pixels were shifted up and down.

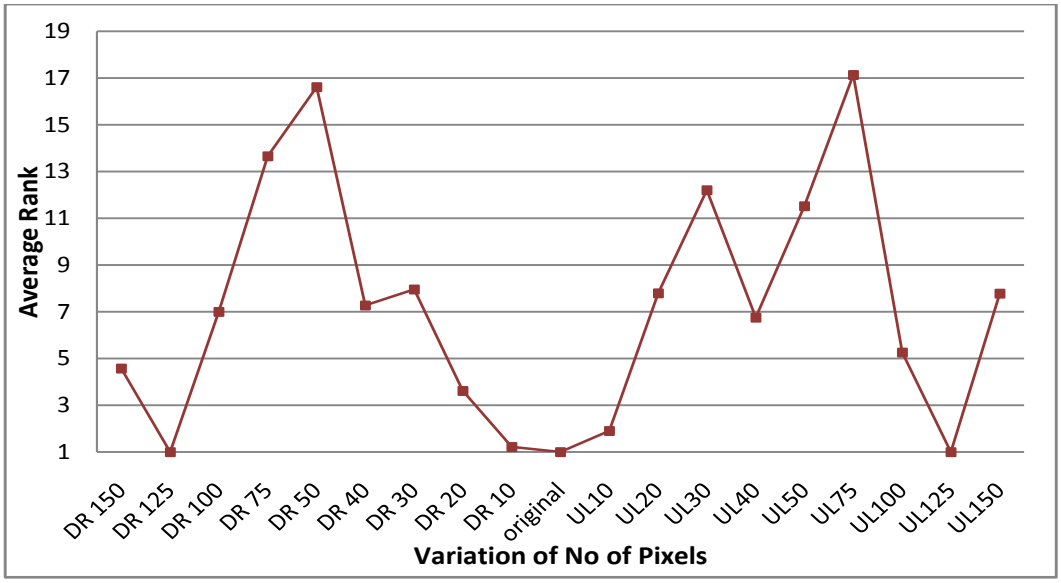


Figure 8.14: Average rank vs image horizontal and vertical shift. Here the pixels were shifted right, left,up and down (D-down, R-right, U-up, L-left).

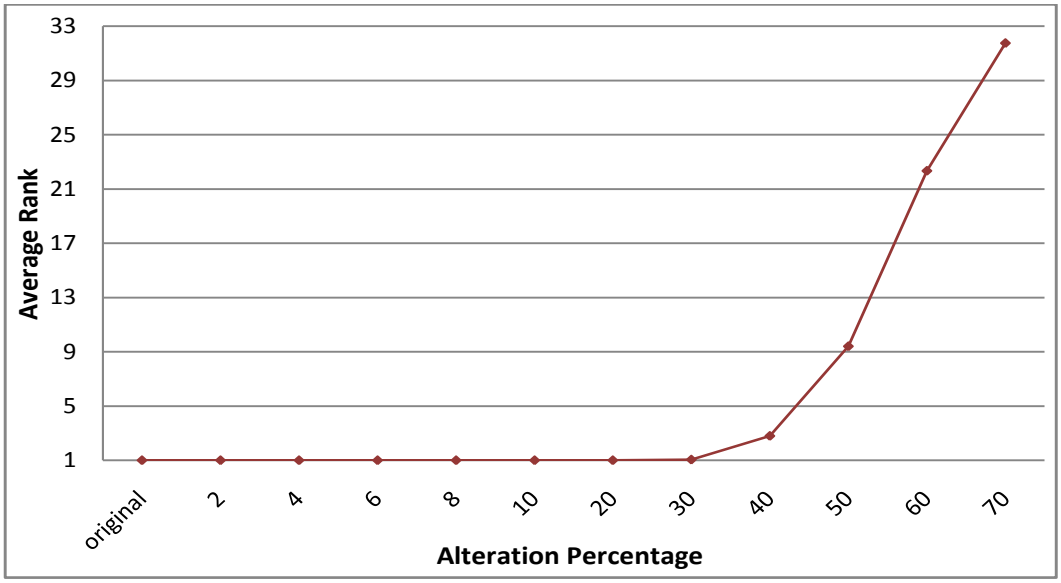


Figure 8.15: Average rank vs image saturation. Decreasing the image saturation in increasing order as a percentage.

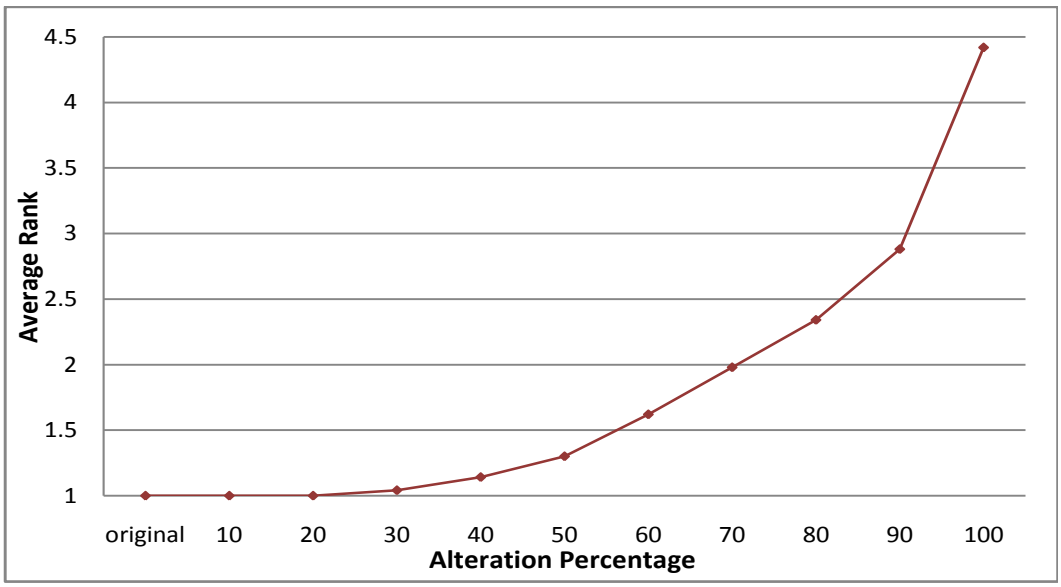


Figure 8.16: Average rank vs increasing image saturation as a percentage.

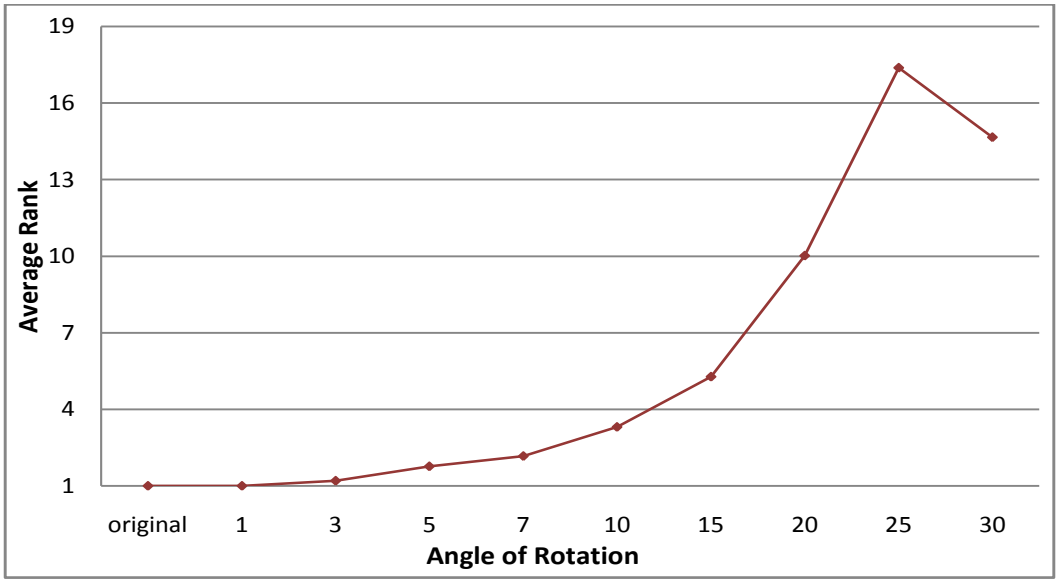


Figure 8.17: Average rank vs image rotation. Rotation is given in degrees.

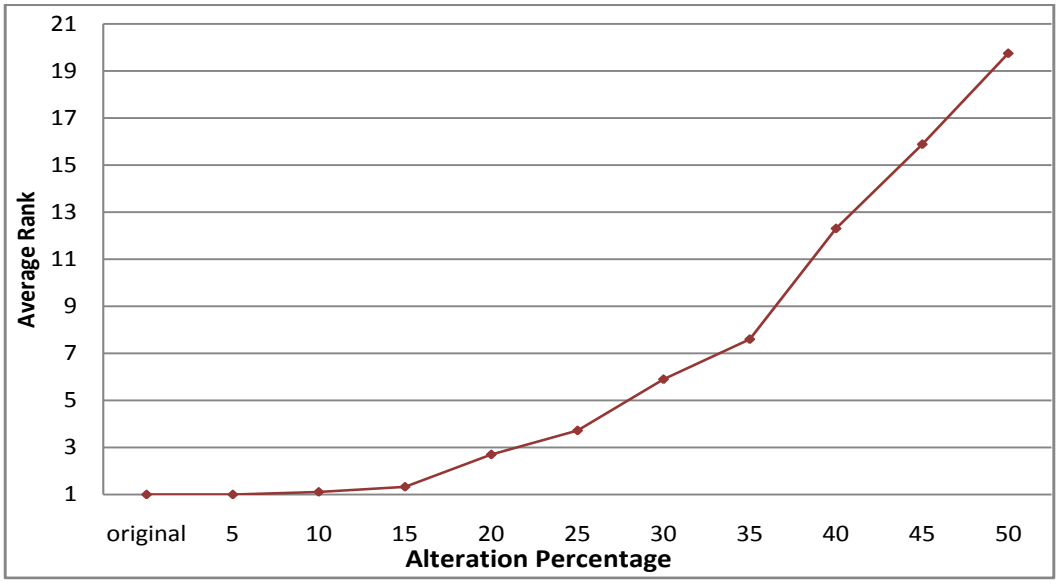


Figure 8.18: Average rank vs image cropping as a percentage. The amount of cropping is given as a percentage.

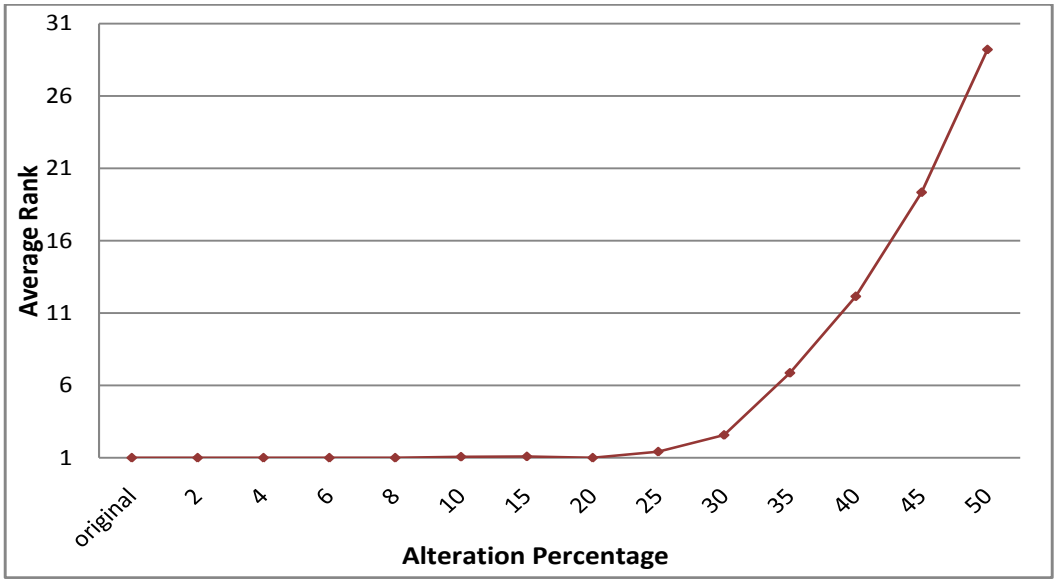


Figure 8.19: Average rank vs image brightness. Brightness is changing as a percentage.

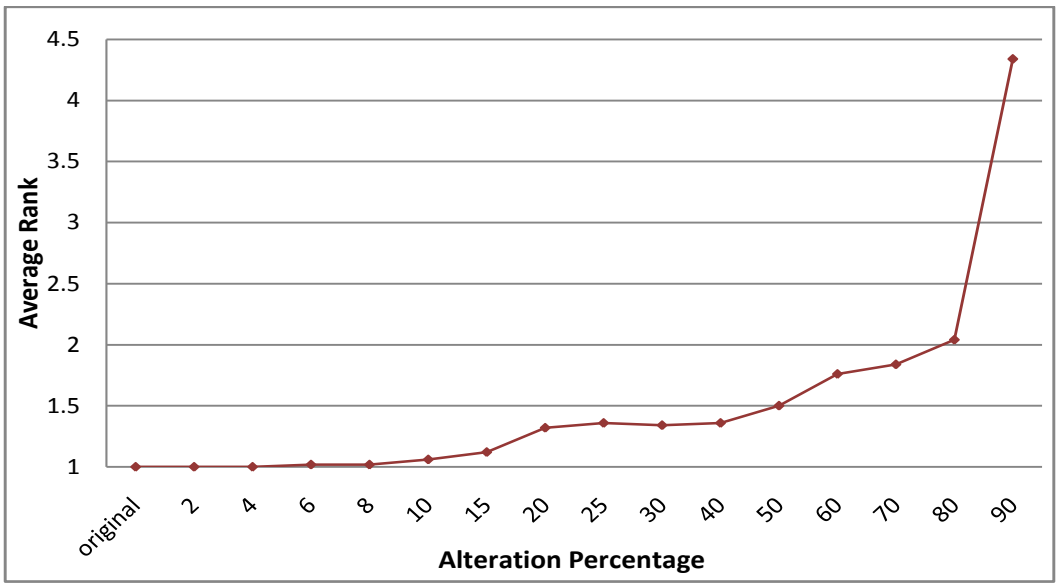


Figure 8.20: Average rank vs image darkness.

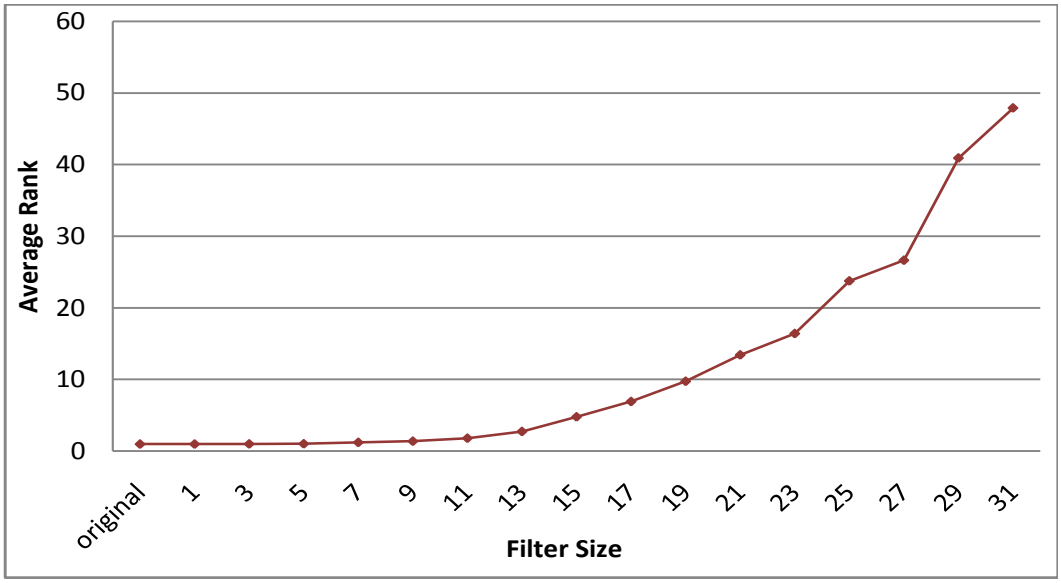


Figure 8.21: Average rank vs Gaussian filter size for blurring. Here $\sigma = 5$.

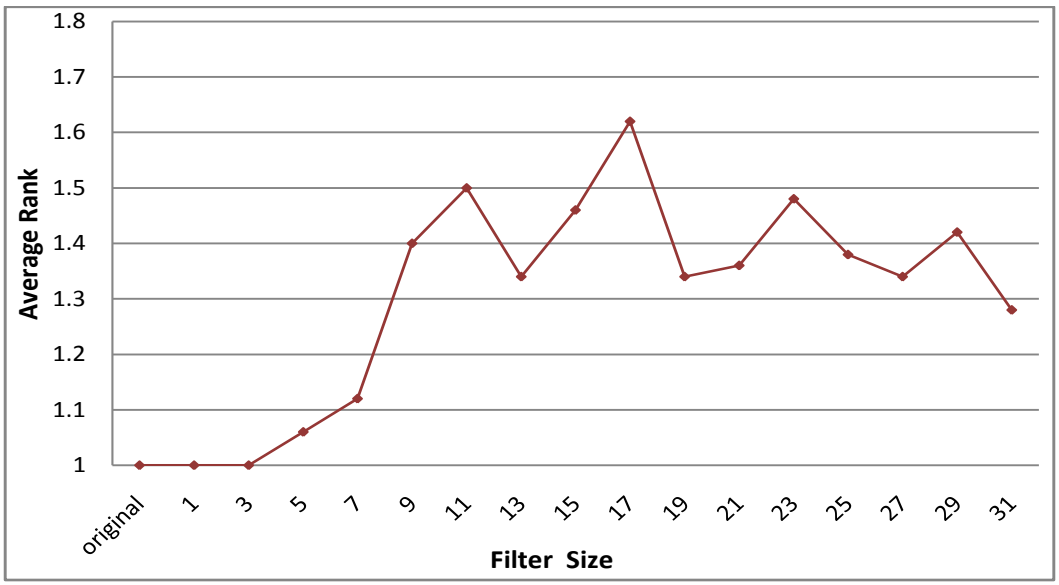


Figure 8.22: Average rank vs Gaussian filter size for sharpening.

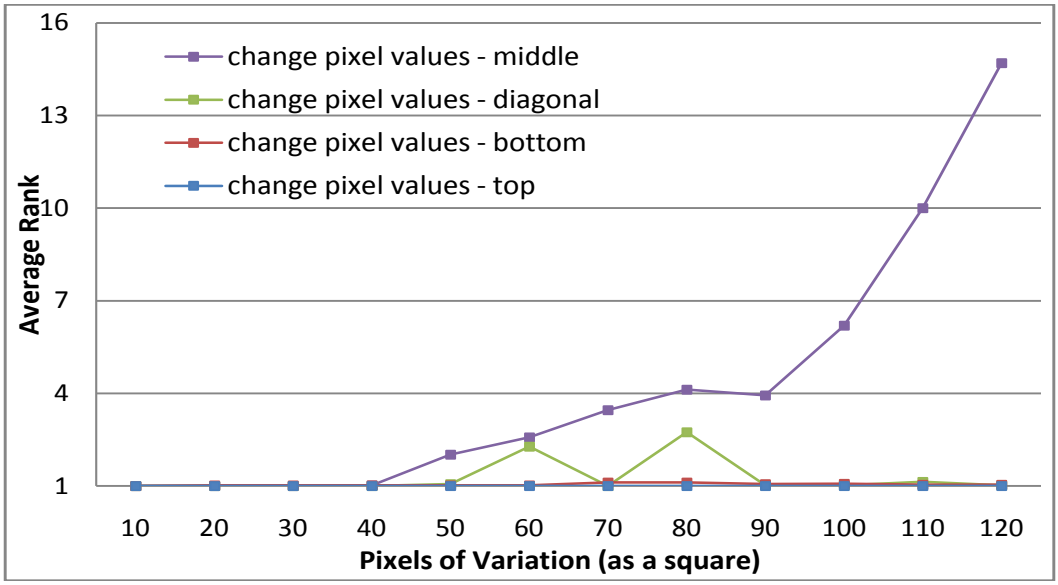


Figure 8.23: Average rank vs shape distortion (single square shape).

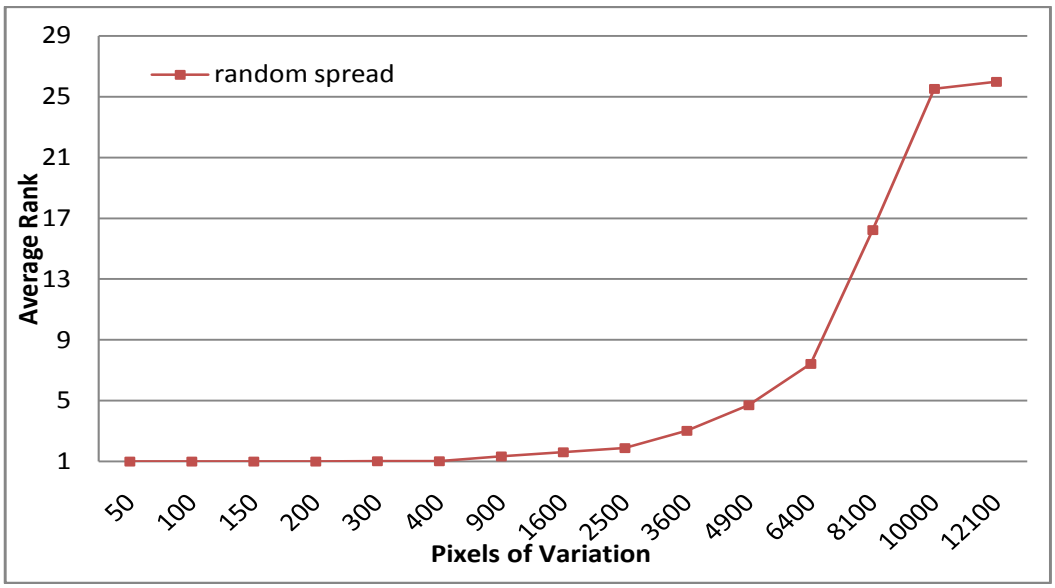


Figure 8.24: Average rank vs shape distortion. Random spread of pixels.

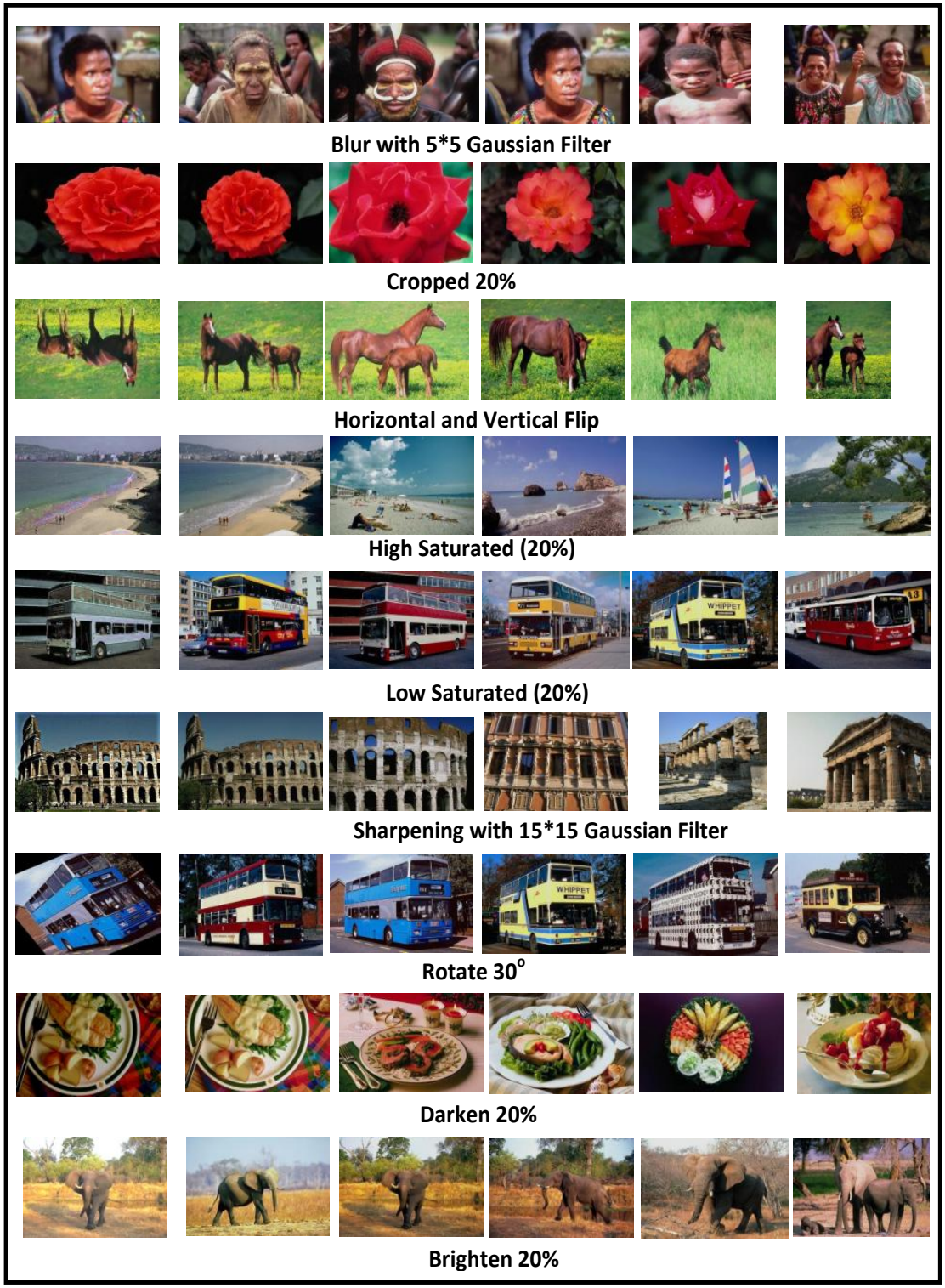


Figure 8.25: The robustness of the proposed approach to image alterations. The first image is the query image and other five are the first five retrieved images.

8.5 Chapter Summary and Conclusions

This dissertation proposed a signature-based image retrieval system as the main contribution, then proposed an RB-RF scheme to further improve the retrieval performance by integrating user perception. This chapter evaluated the scalability of the proposed approaches for efficient and effective CBIR. The initial effectiveness of the system was discussed, with results drawn from previous chapters and we demonstrated that the effectiveness of the system is high. The efficiency of the system is mostly dependent on the database size. However, efficiency is affected in log scale with the increasing database size. Then the efficiency of the system was evaluated by considering searching and feedback time with the varying database sizes and we showed how fast the systems is and how efficient the RB-RF mechanism is and concluded that this can be extended for very large datasets. Feedback searching time does not depend on the database size. Finally, the robustness of the system was evaluated. The system was evaluated over a range of image degradations and we found that the system is robust to most of those alterations under the condition of defined robustness.

Chapter 9

Conclusions and Future Research

Chapter Organisation

This chapter concludes the thesis with a summary of the presented work. Sections 9.1 provides general overview of CBIR and the proposed approaches while Section 9.2 provides the limitation of the research. Section 9.3 summarises the major conclusions and original contributions made in each chapter. Section 9.4 outlines a number of avenues for future research.

9.1 Overview of the Research

The development of the internet and the increased availability of image capturing devices have enabled collections of digital images to grow at a fast pace in recent years and to become more diverse. This has created an ever-growing need for efficient and effective image browsing, searching and retrieval tools. Despite many years of research in this area, an effective general solution has not yet been found. There are several factors that affect overall Content-Based Image Retrieval (CBIR) performance.

The main goal of a CBIR system is to retrieve relevant images to a given query. Therefore, bridging the gap between low-level features and high-level visual semantics is the most important problem that needs to be considered when generating a CBIR system. There are several factors that affect CBIR performance, such as image representation, image descriptors, feature selection and indexing. To achieve

better retrieval performance, all these steps need to carefully considered. However, real user feedback is essential to understanding user perception. mainly because the same user may have a different view on the same image at different times, and different users may have different views on the same image. Therefore, interactive Relevance Feedback (RF) is an essential part of a CBIR system to help narrow down the semantic gap. The efficiency of a CBIR system is also important from the user's point of view, as the elapsed time affects user satisfaction. Therefore, computation complexity must be considered to increase the speed. There must be balance trade-off between efficiency and effectiveness in order to achieve better user satisfaction. Moreover, a CBIR system must be user-friendly that is, that the system can be easily used without having any prior knowledge of image features and the system must not put much of a burden on the user. If all these factors are incorporated, the final outcome is a scalable CBIR system.

Initially, this thesis looked into effective CBIR approaches, techniques and challenges (Chapter 2). After all these studies, we considered on the development of a CBIR system which would achieve better effectivenes. Two CBIR approaches were proposed and their effectiveness was evaluated while studying the most suitable feature combination for CBIR (Chapter 4). Even though the results demonstrated that the proposed systems outperformed when compared with other systems and their suitability for a CBIR task they were not scalable to large datasets. Therefore, it was required to develop a system that could scale to increasingly larger datasets. To this aim, we considered three main research questions. With respect to the first question, (**how can we achieve better retrieval performance in ways analogous to text retrieval by using Bag of Word (BoW) and Random Indexing(RI)?**), we proposed a binary signature-based approach namely, CBIR-ISIG, with the help of the selected feature combination, BoW approach and RI (Chapter 5). The empirical results show that the proposed system offers a trade-off between retrieval quality and speed. The effectiveness of the system was evaluated by comparing it with several systems on different standard datasets and the results demonstrated that the system's ability to provide accurate results as it outperformed the other systems. In addition, a potential application of the CBIR-ISIG system was found for object retrieval by proving its use and effectiveness.

The intention of RF approach motivated this research to in corporate RF into

the proposed CBIR system ,as it helps improve retrieval quality by reducing the semantic gap. Initially, a novel Rank-Based Pseudo RF (RB-PRF) approach was proposed to improve the effectiveness of the CBIR-ISIG system when user interaction is not available, by incorporating the rank order of the initial results rather than giving them equal importance (Chapter 6). This aimed to answer our second research question: **can RF on images be applied in the same way as signature-based RF on text documents to improve retrieval performance?**. In this research work, the feedback process was carried out in the signature space, while considering only a sub-set of results for re-rank which made the RF algorithm computationally efficient. The effectiveness of the RB-PRF demonstrated its effectiveness with CBIR by showing competitive results with the compared approaches. As user interaction is effective in understanding the user perception, explicit user feedback was incorporated with the CBIR-ISIG system by providing the user a chance to refine the query (Chapter 7). The user-engaged experiments demonstrated the difference between classification and user requirement while showing significant improvement in retrieval quality.

Our third research question was: **how can we achieve fast retrieval results and scale a CBIR approach to a very large image database?** The aim of this question was to develop a scalable system that could operate on larger datasets. This question was partially answered by the first two questions and the results from the experiments demonstrated the effectiveness and efficiency of the proposed approaches and the ability for the system to be applied to larger datasets. In addition, results from the robustness experiments imply that the CBIR-ISIG system is robust to a large range of alterations.

It is hoped that this dissertation provides important contributions to the research detailed in Section 9.3

9.2 Limitations

There were some limitations when developing and evaluating the proposed approaches.

There are a plethora of features available for feature extraction in CBIR and we did not study all those features when we selected features for this research. This research used some the best and fast colour, texture and shape retrieval techniques

together in order to address general images. Therefore, there may be other feature combinations which provide effective retrieval performance.

In addition, we used a scaling factor for RF and in that, the term w is a decay factor, which determines how fast the feedback from images decays with rank. This value works well over a wide range of collections and experiments. We selected w values after experimenting with PRF and simulated RF. However, we did not experiment with real users as it is a cumbersome task for users.

The run-times recorded in our experiments could be contextualised and contrasted with results reported in prior work. However, a direct comparison is not possible because the reported results were obtained using different hardware architecture and experimental settings and the implementations of previously proposed methods are not freely available. Therefore, we speculated by considering system architecture and the size of the datasets used.

9.3 Summary of Contributions

A number of original contributions have been made and presented in this dissertation as follows:

- **The development of a Content-Based Image Retrieval system based on binary Image SIGnatures (CBIR-ISIG) in order to achieve accurate results efficiently:** Initially, this work focusedon identifying various sizes of semantic image feature building blocks which can be used to represent an image as a bag of semantic image features. Then the grid-based image decomposition, modified BoW representation (symbolical representation) and RI provided effective and efficient binary image signatures to search the image database. Moreover, this provides solutions to the problem of high computational complexity of indexing, retrieval process, memory and disk space requirements with the use of binary image signatures. The retrieval performance outperformed the existing systems in the literature based on different benchmark datasets which illustrates that the proposed approach has a high potential to retrieve correct images. The system demonstrated significant effectiveness and efficiency which, in turn, can be extended to a large collection. This CBIR system is presented in Chapter 5.

- **The development of Rotational Invariant Bag-of-visual Words representation (RIBoW):** With the motivation of binary signature-based CBIR, we extended this work for the application of object retrieval. The incorporation of circular image decomposition and circular shift with binary image signature presentation provided an effective rotation invariant representation which can be used in CBIR, especially in object retrieval. The results from the experiments highlighted its effectiveness and robustness for rotation invariant image retrieval by transcending retrieval effectiveness when compared with other BoW approaches in object retrieval on the benchmark dataset of objects. Therefore, this approach can be used in object retrieval applications to gain effective and efficient retrieval performance. This RI-BoW approach is presented in Section 5.5.
- **The development of relevance feedback techniques that incorporate rank information which allows images to be retrieved effectively and efficiently:** PRF has proven to be an effective mechanism in improving retrieval accuracy. This is very useful, especially in the absence of an actual user to provide feedback. An original, simple yet effective RB-PRF that takes into account the initial rank order of each image to improve retrieval accuracy is proposed. This RB-PRF mechanism innovates by making use of binary image signatures to improve retrieval precision by promoting images similar to highly-ranked images and demoting images similar to lower-ranked images. This provides a mechanism to perform RF in the signature space without ever going back to the original image features. Furthermore, the initial list of signatures (only a sub-set of the first-pass results) that are being re-ranked is already in memory following the initial search process, so the process is computationally efficient. Empirical evaluations based on several standard benchmarks demonstrated the effectiveness of the proposed RB-PRF mechanism to CBIR by showing competitive results with the compared approaches. This RB-PRF approach is presented in Chapter 6.

  Then the RB-PRF approach was extended to explicit user feedback to improve retrieval by understanding the user's perception while providing the user with an opportunity to refine the query. However, most proposed algo-

rithms neglect experimental evaluation with real users. This research work presents a methodology for the use of binary signature-based image retrieval with a user in the loop to improve image retrieval performance. The significance of this study is twofold. First, it shows how to effectively use explicit relevance feedback with signature-based image retrieval to improve retrieval quality. Second, this approach provides a mechanism for end-users to refine their image queries. Unlike text retrieval systems where users are able, and generally prefer, to reformulate their text queries to improve search results, there is no effective way to reformulate an image query. This approach provides a solution to this problem. Extensive experiments were carried out to study the behaviour and optimal parameter settings of this approach. Empirical evaluations based on various standard benchmarks demonstrated the effectiveness of the proposed approach in improving the performance of CBIR. This explicit RB-RF approach is presented in Chapter 7.

- **The scalability evaluation of the CBIR-ISIG system and RB-RF approaches in terms of effectiveness, efficiency and robustness in retrieval:** Initially, the effectiveness of the CBIR-ISIG system and the proposed RB-RF approaches with PRF and simulated user feedback, and explicit user feedback was evaluated and then the efficiency of feature extraction and retrieval of images at first pass retrieval and RF was evaluated. The robustness was evaluated for a range of alterations to understand the CBIR-ISIG system's robustness to different degradations of images. With all the results from these empirical evaluations it was concluded that using the proposed CBIR-ISIG system and RF approaches in CBIR significantly improves retrieval effectiveness and efficiency. Moreover, this signature-based representation is robust to a range of image degradations. This scalability evaluation is presented in Chapter 8.

In this dissertation, several CBIR methods are introduced, specifically a signature-based approach (CBIR-ISIG) for efficient and effective general image retrieval. An RF approach is proposed to improve retrieval performance further and the system was evaluated for the scalability. Therefore, the main challenges are solved by introducing different techniques. Despite these contributions, there

are many opportunities to achieve improved results. The following section outlines different options that can be used to develop the proposed approaches in future work.

9.4 Future Work

A number of possible avenues for future research have been identified. The proposed approaches can be enhanced in terms of efficiency, effectiveness and robustness.

The main drawback of the proposed CBIR-ISIG system is that it consumes considerable time for pre-processing. To solve this issue, parallel processing can be applied to pre-processing in retrieval system when the query image from outside the database to increase its efficiency. Moreover, experiments must be carried out to find the effect of feature extraction in compressed domain as feature extraction in compressed domain will help to increase efficiency.

More information of the type of users, user queries and studying the users' views may help to gain greater accuracy in evaluation and lead to accurate retrieval based on the user's view. However, this is very time-consuming and may be a cumbersome task for the user. Therefore, controlled user experiments must be arranged. In this work, experiments will be controlled by giving users a subset of images to provide their feedback at their convenience and these images will be same, by giving users exactly the same setting, same amount of time, same instructions and the demographic of users will be known. Furthermore, this experiment will be lab-based similarly to the experiments, which were done in Chapter 7. In addition, developing the system as a web-based system will provide an opportunity to further study user behaviour and collect more data which can then be used for simulated experiments to study leads and drawbacks. However, this will be done by sending a link to selected users to provide feedback after they confirm their participation and same setting will be used as lab-based. Therefore, experimental setting is under control. The differences between the lab-based and web-based are, in lab-based one can monitor how long it takes to complete a task, gives users better instructions, helps them in difficulties, resolves misunderstandings, while in web-based more user data can be collected and convenient for users . Therefore, lab-based system have more control over the setting than web-based. The use of a combination

of concept-based and content-based approaches has been suggested in order to improve retrieval performance in terms of effectiveness.

Focused relevance feedback is an RF approach which takes feedback in the form of highlighting the relevant area. The CBIR system then incorporates that information to re-rank the rest of the collection. This will be useful in acquiring higher precision and we expect that this will improve results more than by using full image feedback. In addition, this will be helpful to the user as they can collect the required information at the early stage without them having to put in much effort. Therefore, focus relevance feedback can be incorporated with CBIR-ISIG in order to improve effectiveness.

9.5 Final Remarks

The proposed CBIR-ISIG system is significant as it provides significant effectiveness, efficiency and robustness in CBIR. These features enable the CBIR-ISIG system to be extended for large databases and applications such as object retrieval. The CBIR-ISIG, when incorporated with RB-RF approaches, is significant because RF performs in the signature space and helps achieve competitive performance efficiently by re-ranking only the sub-set (initial list) of the first-pass results. Moreover, explicit relevance feedback provides a mechanism for end-users to refine their image queries. Therefore, it can be argued that CBIR-ISIG with RB-RF provides an effective, efficient and robust system for CBIR.

Appendices

Appendix A

User Interface of the CBIR-ISIG System

This Chapter gives an overview of the user interface of the CBIR-ISIG system.

Figure A.1 gives the initial appearance of the CBIR-ISIG system. It is very simple and it has two options. If user wants just search then user has to upload an image and click search button. If user requires results based on specific feature, he/she can select that option from the top right hand corner. figure A.2 shows an example for that by showing options for colour. User can selected any feature from the colour feature as listed. There is a option to selct image type is required as shown in A.3.

When the searching process starts user first has to upload an image. When user click "Upload" it directed to the database as shown in A.4 and when user click on an image it appears in the interface and results are shown after click on "Search" as shown in A.5.

After seeing the given results user can give their feedback to the system and get new results list back. Figure A.6, A.7 and A.8 shows the given user FB of different users. In evaluation phase user was asked to run the FB search for five iterations. Each time user has to click on "Feedback & Search" button to receive FB results. After fifth iteration user is given a message by notifying that five iterations are over ready for the next step as shown in figure A.9. In there user is asked to click "First Pass" button which was inactive during the FB iterations. This was done

because the user may confuse when several buttons are active.

After user click on ”First Pass” user is guided by the system to select relevant images from given 100 images as shown in figure A.10. However, these 100 images are shown by five separate pages which user can see them by click on ”Next” and ”Previous” buttons which activates after click on ”OK” on the shown message. Figure A.11 shows intermediate results page of the first pass retrieval results, ”Next” and ”Previous” are activate and ”Feedback & Search” is inactive here. After user comes to the last page of the first pass results ”Feedback & Search” button is active to be clicked on as shown in figure A.12.

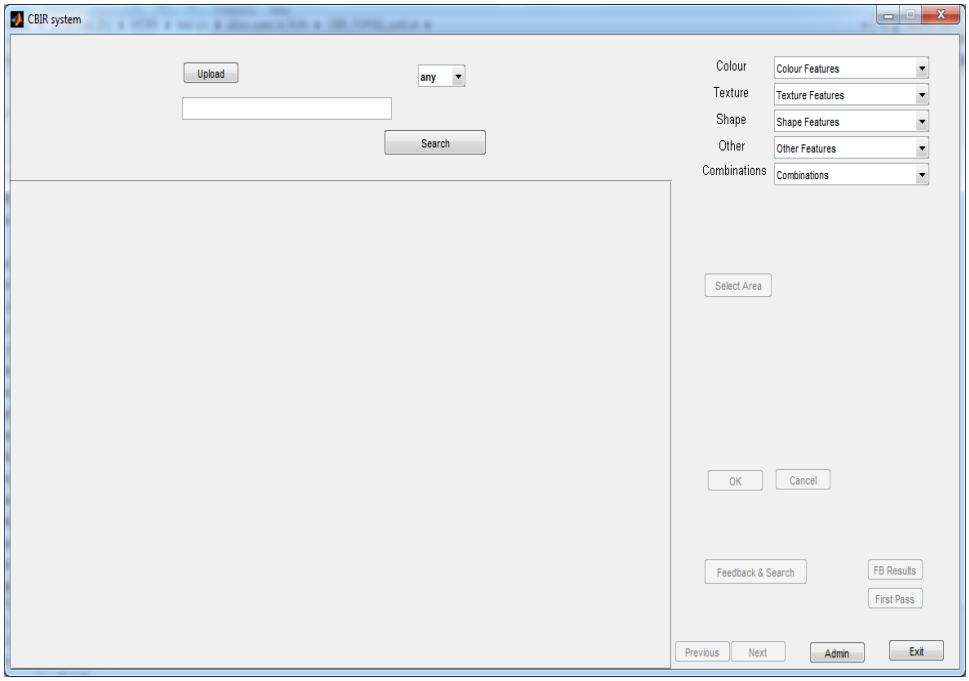


Figure A.1: Initial look of the GUI.

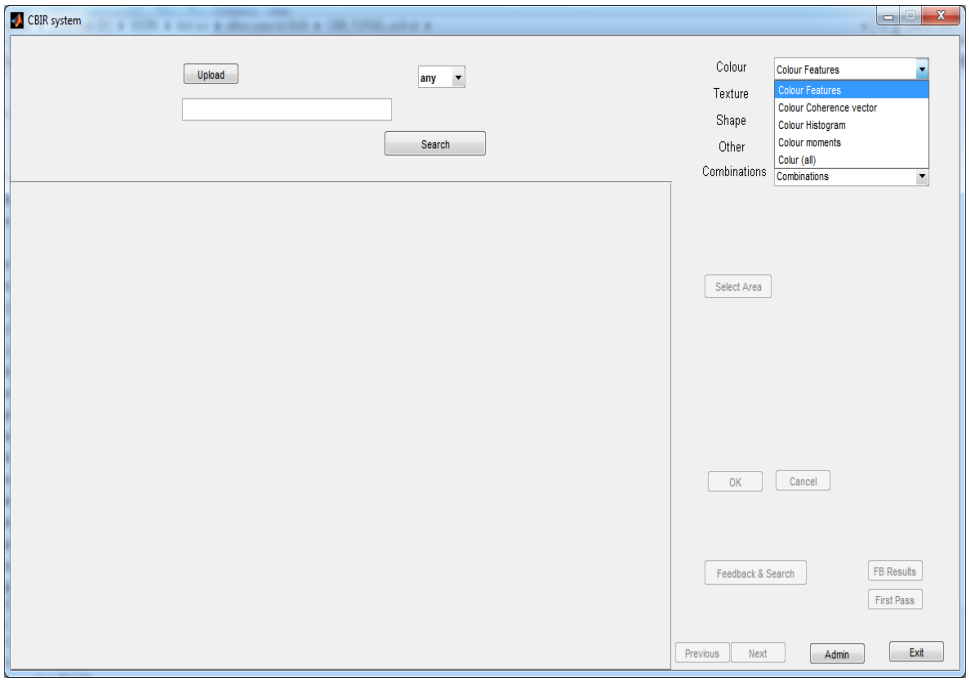


Figure A.2: User is allowed select preferred feature if they need. Here listed the features for colour.

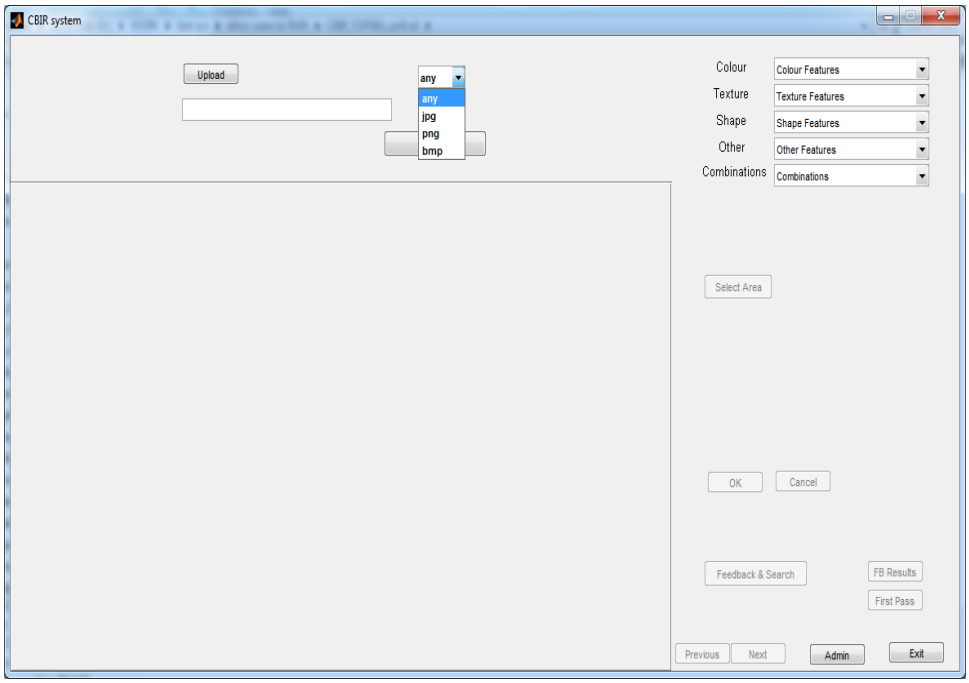


Figure A.3: Different types of images can be selected.

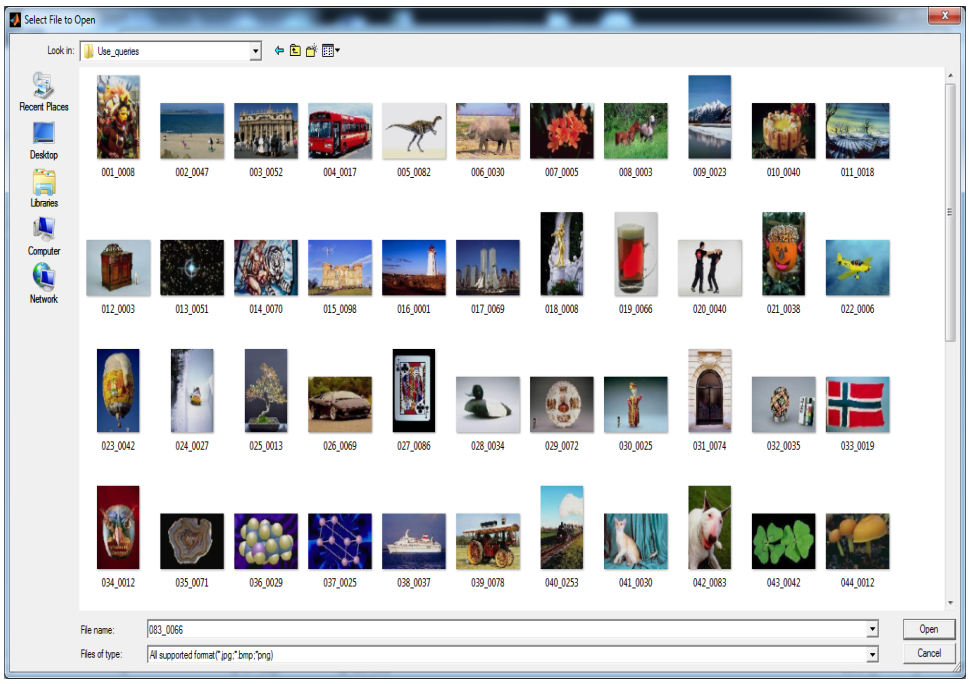


Figure A.4: Image database is shown when user click Upload button.

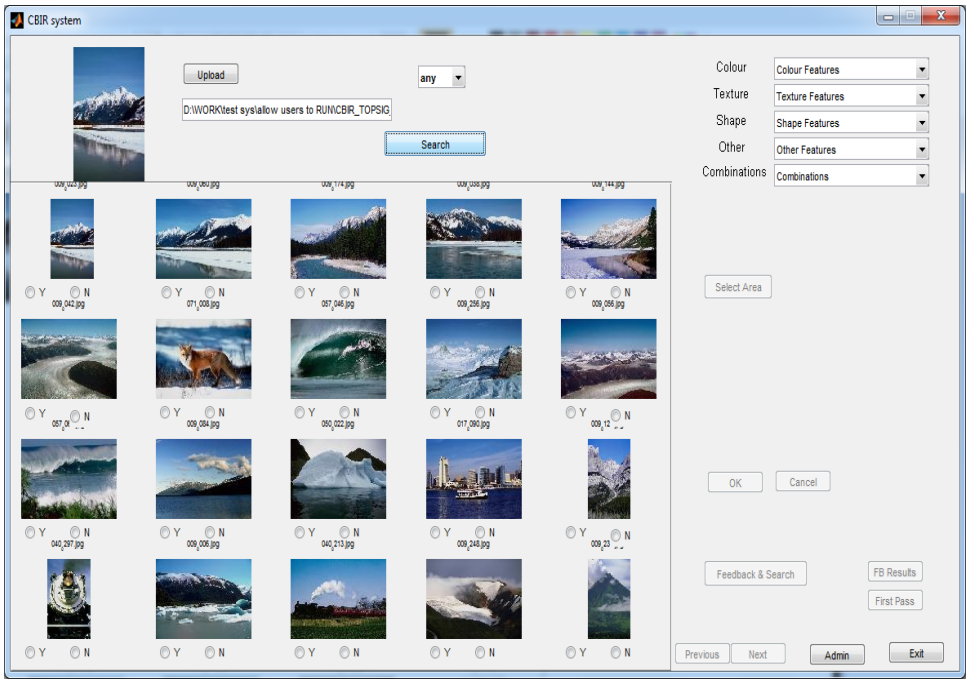


Figure A.5: Retrieval results are displayed after clicking on Search Button.

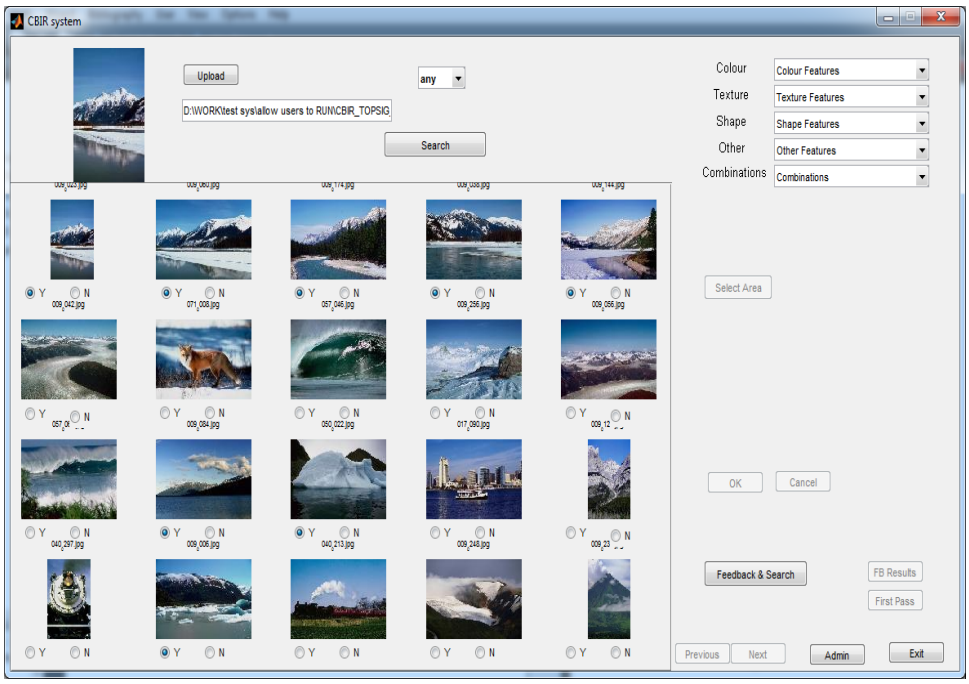


Figure A.6: User FB for the given query (User1).

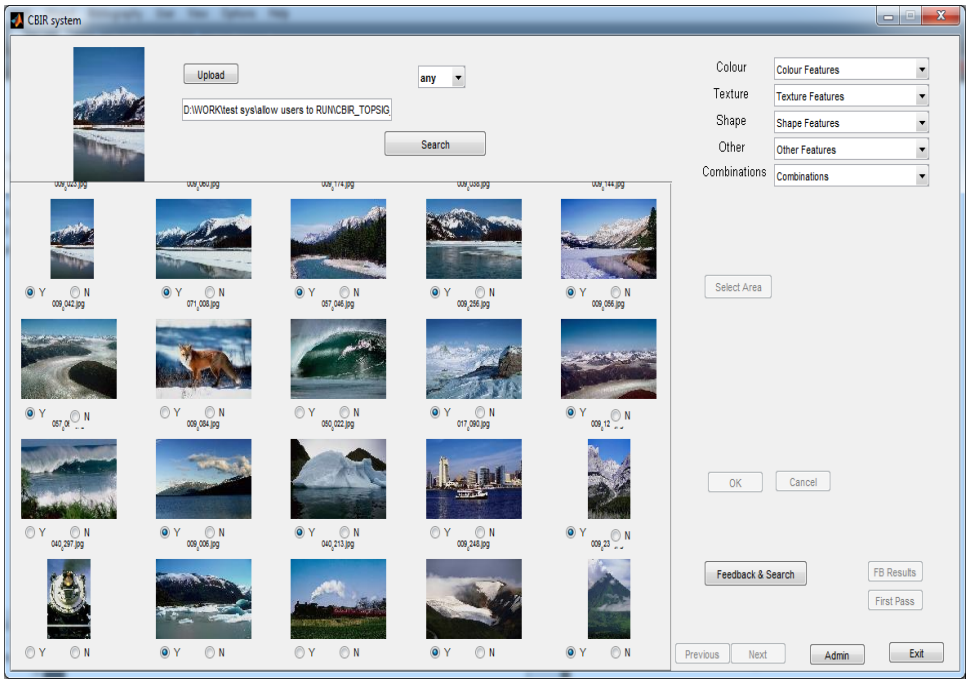


Figure A.7: User Feedback for the given query (User2).

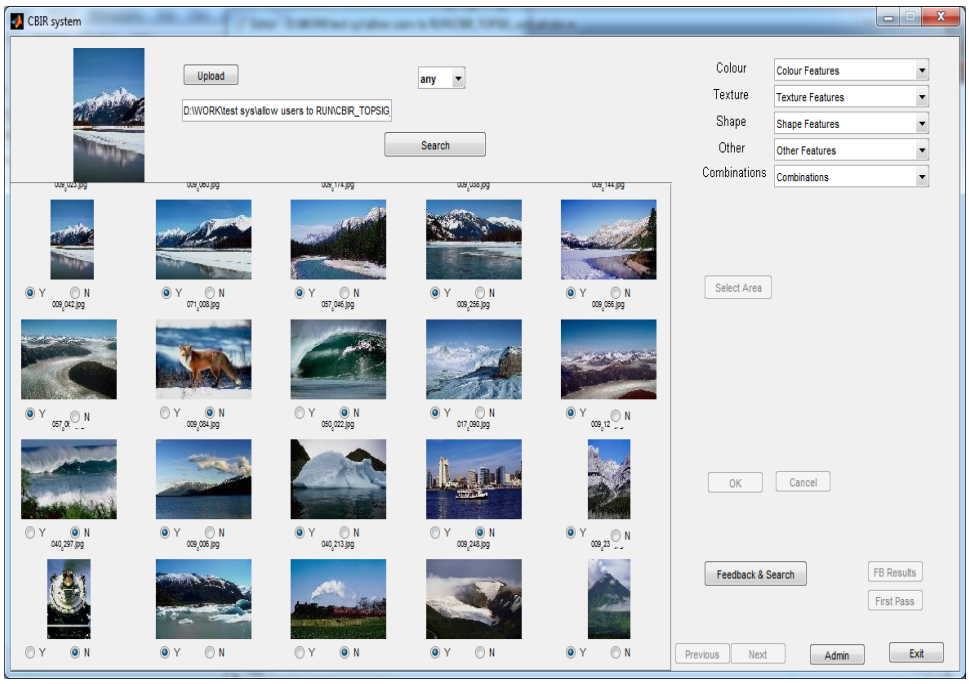


Figure A.8: User Feedback for the given query (User3).

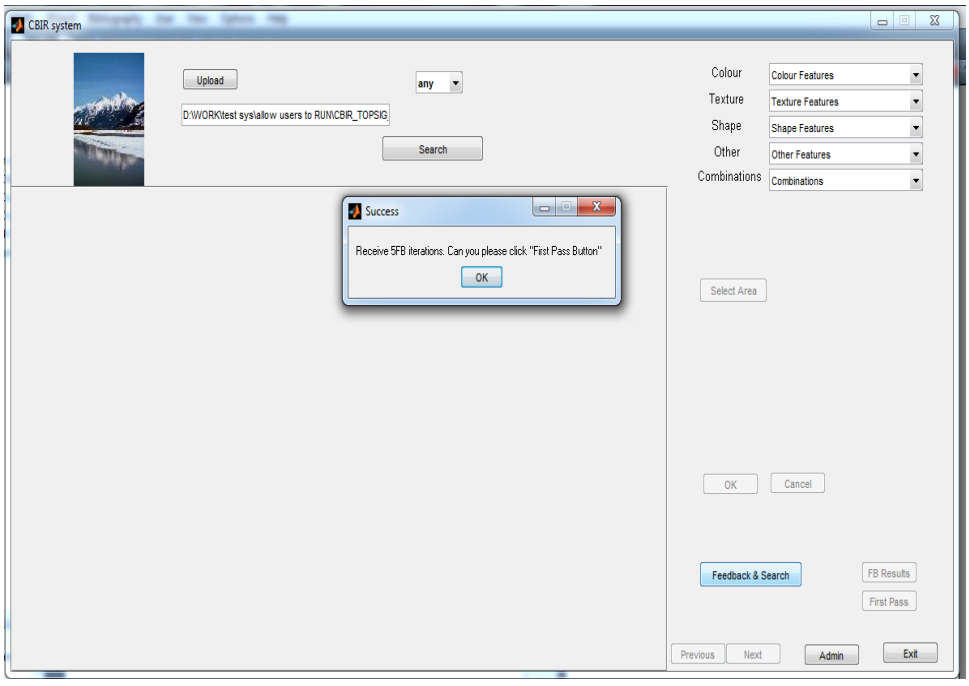


Figure A.9: After five iterations.

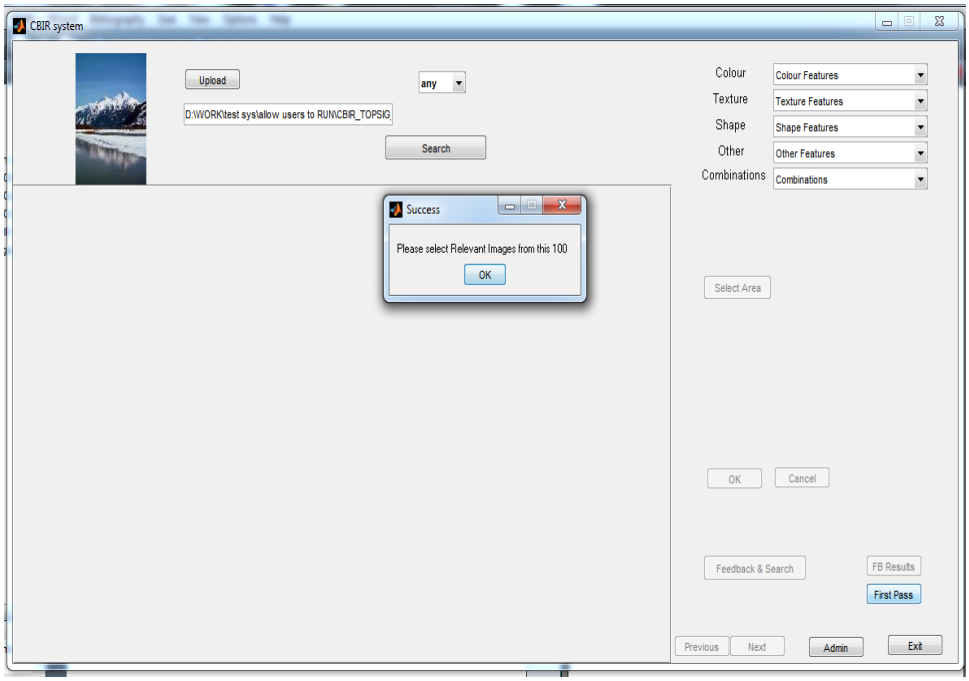


Figure A.10: First Pass clicked.

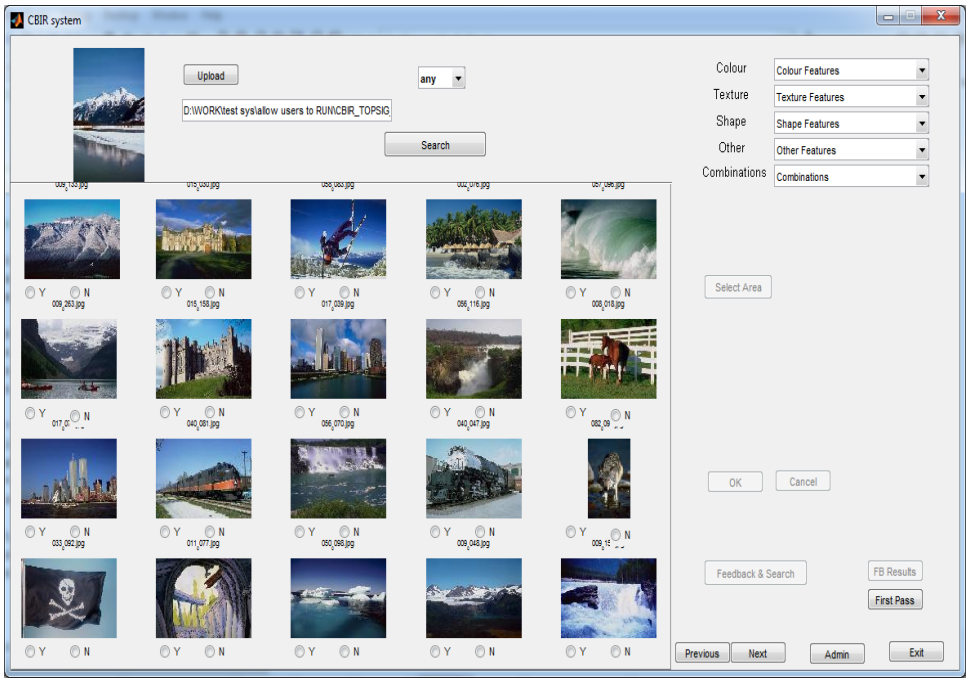


Figure A.11: Intermediate results page after clicking on First Pass. User can see 100 images using Previous and Next Button

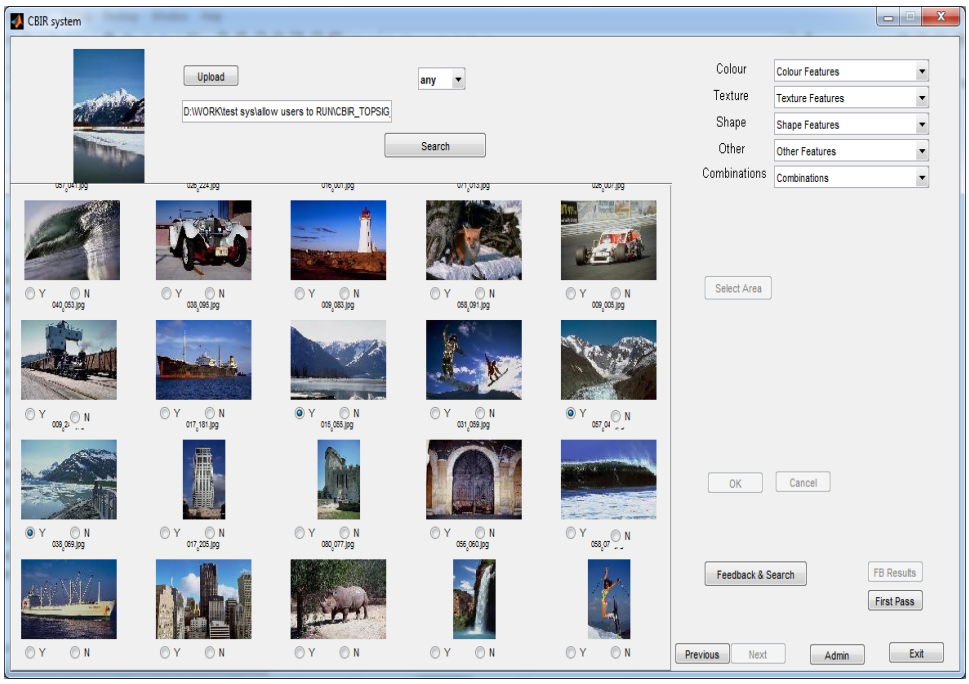


Figure A.12: Last page of the first pass results for the given query.

Usability

The success of efficient interaction between a user and a CBIR system depends on system's usability. When the usability of a system is considered, it concerns not only the user interface but several other factors can be crucial for system's performance in general. These usability issues and how CBIR-ISIG system handles these issues as listed below.

- Query specification : the system should allow flexibility of formation and modification of search request.
  - CBIR-ISIG system allows user to change their query if required and every time they only have to click "Upload" button when they want to alter.
- Results presentation : the system should provide user results in a clearly arranged presentation in a user interface.
  - CBIR-ISIG system shows 20 images according to their similarity, at a time as user always expect images with highest similarity and sometime user may be confused by seeing large amount of images. This can be seen from the listed GUI figures here.
- Retrieval speed : the system should have efficient searching mechanism to give results without allowing user to wait.
  - CBIR-ISIG system use binary image representation as final representation which is used in searching procedure. Hamming distance is used as similarity measure in the Topsig signature based search system, for searching. CBIR-ISIG system is fast because it inherits properties of the Topsig. This can be seen from figure 8.6 and 8.7.
- Relevance feedback : the system should provide effective relevance feedback system to use.

- CBIR-ISIG system provides effective FB mechanism. Because simple binary feedback is used, either ”yes” or ”no” is required for the images which needs to be provided by the user. The proposed RB-RF approach works quite well on CBIR with improved results. This can be seen from results in Chapter 6and Chapter 7.

- Interaction period : The system should be able to react instantaneously.

  - CBIR-ISIG system provide an efficient RF searching system which needs only milliseconds to search and present irrespective of the size of the dataset. This can be seen from figure 8.11.

- Easiness to use : the system should be easy to learn, efficient and pleasant to use.

  - CBIR-ISIG system is easy to use and no prior knowledge is required and it has law error rate to attract and satisfy the user. Moreover user is guided by the system.

Usability of the system was discussed here. However, the usability was not evaluated with the real users.

Appendix B

Interactive Relevance Feedback Evaluation : Ethics Clearance

The project is concerned with a search engine for images. Unlike a text search engine, which is using keywords or text questions to find relevant documents, this system is using example images as queries. After conducting an initial search, the system presents several images, deemed to be similar in some sense to the query image. The user is then able to indicate if, and which, of the returned images are relevant images, by clicking on the relevant images. The system is then able to refine its search and try and present an improved set of result images, using the feedback. This process can be repeated as many times as the user wishes to. Therefore, human interaction is essential. Human ethics clearance was obtained to carry out the evaluation of relevance feedback, which was is exempted from ethical review by the QUT Human Research Ethics Committee (Exemption number:1600000622). The evaluation of the system is done in two sessions rather than having one session. Nevertheless, the evaluation process and the requirements have remained the same.

Content-Based Image Retrieval with Interactive Relevance Feedback

PARTICIPANT INFORMATION FOR QUT RESEARCH PROJECT

QUT Ethics Exemption Number 1600000622

RESEARCH TEAM

Principal Researcher: Nanayakkara Wasam Uluwitige Dinesha Chathurani, PhD Student

Associate Researchers: Prof. Shlomo Geva, Prof. Vinod Chandran (QUT) and Prof. Wageeh Boles.

DESCRIPTION

This project is being undertaken as part of the PhD thesis of Nanayakkara Wasam Uluwitige Dinesha Chathurani, which is focused on improving content based image retrieval results through user feedback. Specifically, the aim of this thesis is to investigate user perception images rather than classification.

The research team requests your assistance to evaluate the relevance feedback approach on the content-based image retrieval system .

PARTICIPATION

The researchers are looking for QUT students and staff who are willing to give feedback on the content-based image retrieval system.

Participation in this research is completely voluntary, and if you do agree to participate, you are free to withdraw from participation at any time without comment or penalty. Your decision to participate, or not participate, will in no way impact upon your current or future relationship with QUT (e.g. your grades).

You will be asked to use the system at an agreed location that is convenient to you. The experiment will take approximately 45 minutes of your time for the first session and 30 minutes for the second session.

During the first sessions, you will be asked to interact with the system as follows:

1 You will be shown a query image.

2 You will be shown 20 images that the search engine had identified as potentially relevant with respect to the query image.

3 You will be required to click any image that you deem to be relevant. Relevance can be determined according to your subjective assessment and you do not need to explain. Just click relevant images.

4 When you are done the system will refresh the screen with 20 new images. It will use your feedback from the previous step in an attempt to find more relevant images.

5 Steps 2 to 4 will be repeated until you have viewed 5 screens of results.

6 You will be asked to select relevant images from shown 100 images.

7 A new query image will be presented and steps 2 to 6 repeated.

8 The experiment will end after 80 query images were processed.

9 You may end the experiment earlier if you wish. You may take breaks as you wish.

During the second sessions, you will be asked to interact with the system as follows:

1 You will be shown a query image.

2 You will be shown a image that the search engine had identified as potentially the best relevant with respect to the query image.

3 You will be required to click yes if you feel that the full image is relevant and select a area of the image that you deem to be relevant or click no otherwise. Relevance can be determined according to your subjective assessment and you do not need to explain.

3 When you are done the system will refresh the screen with a new image. It will use your feedback from the previous step in an attempt to find more relevant images.

4 Steps 2 to 3 will be repeated until you have viewed 50 images.

5 A new query image will be presented and steps 2 to 4 repeated.

6 The experiment will end after 10 query images were processed.

7 You may end the experiment earlier if you wish. You may take breaks as you wish.

EXPECTED BENEFITS

It is expected that this project will not benefit you directly. However, it may benefit all the users who do search images using query by example. The research team seeks to benefit from this experiment by studying the utility of user feedback in improving the search results of image retrieval systems.

RISKS

There are minimal risks associated with your participation in this project. These include your inconvenience and time. Therefore, we ask you to inform us about any convenient time during the given duration and if you are not forced attend if you do not have enough time to participate.

We believe there are minimal risks with your participation in this feedback, which you should consider:

No any detail will be taken from you and you can attend at any convenient time.

PRIVACY AND CONFIDENTIALITY

No personal details or image will be taken.

CONSENT TO PARTICIPATE

Due to the nature of the project a verbal consent mechanism will be used.

CONCERNS/COMPLAINTS REGARDING THE CONDUCT OF THE PROJECT

QUT is committed to research integrity and the ethical conduct of research projects. However, if you do have any concerns or complaints about the ethical conduct of the project you may contact the QUT Research Ethics Unit on 3138 5123 or email ethicscontact@qut.edu.au. The QUT Research Ethics Unit is not connected with the research project and can facilitate a resolution to your concern in an impartial manner.

Bibliography

[Abubacker and Indumathi, 2010] Abubacker, K. and Indumathi, L. (2010). Attribute associated image retrieval and similarity reranking. In *Communication and Computational Intelligence (INCOCCI), 2010 International Conference on*, pages 235–240.

[Agarwal et al., 2013] Agarwal, S., Verma, A., and Singh, P. (2013). Content based image retrieval using discrete wavelet transform and edge histogram descriptor. In *Information Systems and Computer Networks (ISCON), 2013 International Conference on*, pages 19–23.

[Aman et al., 2010] Aman, J., Yao, J., and Summers, R. (2010). Content-based image retrieval on ct colonography using rotation and scale invariant features and bag-of-words model. In *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pages 1357–1360.

[Amanatiadis et al., 2009] Amanatiadis, A., Kaburlasos, V., Gasteratos, A., and Papadakis, S. (2009). A comparative study of invariant descriptors for shape retrieval. In *Imaging Systems and Techniques, 2009. IST '09. IEEE International Workshop on*, pages 391–394.

[Amanatiadis et al., 2011] Amanatiadis, A., Kaburlasos, V., Gasteratos, A., and Papadakis, S. (2011). Evaluation of shape descriptors for shape-based image retrieval. *Image Processing, IET*, 5(5):493–499.

[Arampatzis and Kamps, 2009] Arampatzis, A. and Kamps, J. (2009). A signal-to-noise approach to score normalization. In *ACM International Conference on Information and Knowledge Management CIKM*, page 797Ű806.

[arszalek and Schmid, 2006] arszalek, M. M. and Schmid, C. (2006). Spatial weighting for bag-of-feature. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Aslam and Montague, 2001] Aslam, J. A. and Montague, M. (2001). Models for metasearch. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '01, pages 276–284, New York, NY, USA. ACM.

[Azzopardi, 2011] Azzopardi, L. (2011). The economics in interactive information retrieval. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 15–24. ACM.

[Azzopardi and Zuccon, 2015] Azzopardi, L. and Zuccon, G. (2015). Building and using models of information seeking, search and retrieval: Full day tutorial. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, pages 1107–1110, New York, NY, USA. ACM.

[Banda et al., 2013] Banda, J., Angryk, R., and Martens, P. (2013). On dimensionality reduction for indexing and retrieval of large-scale solar image data. *Solar Physics*, 283(1):113–141.

[Banerjee et al., 2009] Banerjee, M., Kundu, M. K., and Maji, P. (2009). Content-based image retrieval using visually significant point features. *Fuzzy Sets and Systems*, 160(23):3323 – 3341. Theme: Computer Science.

[Belongie et al., 1998] Belongie, S., Carson, C., Greenspan, H., and Malik, J. (1998). Color- and texture-based image segmentation using em and its application to content-based image retrieval. In *Computer Vision, 1998. Sixth International Conference on*, pages 675–682.

[Bian and Tao, 2010] Bian, W. and Tao, D. (2010). Biased discriminant euclidean embedding for content-based image retrieval. *Image Processing, IEEE Transactions on*, 19(2):545–554.

[Bucak et al., 2011] Bucak, S. S., Jin, R., and Jain, A. K. (2011). Multi-label learning with incomplete class assignments. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2801–2808.

[Cai et al., 2004] Cai, D., He, X., Li, Z., Ma, W.-Y., and Wen, J.-R. (2004). Hierarchical clustering of www image search results using visual, textual and link information. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, MULTIMEDIA '04, pages 952–959, New York, NY, USA. ACM.

[Carson et al., 1997] Carson, C., Belongie, S., Greenspan, H., and Malik, J. (1997). Region-based image querying. In *Content-Based Access of Image and Video Libraries, 1997. Proceedings. IEEE Workshop on*, pages 42–49.

[Carson et al., 2002] Carson, C., Belongie, S., Greenspan, H., and Malik, J. (2002). Blobworld: image segmentation using expectation-maximization and its application to image querying. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(8):1026–1038.

[Chang and Fu, 1980] Chang, N.-S. and Fu, K.-S. (1980). Query-by-pictorial-example. *IEEE Transactions on Software Engineering*, SE-6(6):519–524.

[Chappell, 2015] Chappell, T. (2015). *Scalable Document Hashing and Retrieval*. PhD thesis, Science and Engineering Faculty of Queensland University of Technology.

[Chappell and Geva, 2015] Chappell, T. and Geva, S. (2015). Topsig: A scalable system for hashing and retrieving document signatures. In *Information Retrieval Technology*, pages 447–452. Springer.

[Chappell et al., 2013] Chappell, T., Geva, S., Nguyen, A., and Zuccon, G. (2013). Efficient top-k retrieval with signatures. In *Proceedings of the 18th Australasian Document Computing Symposium*, ADCS '13, pages 10–17, New York, NY, USA. ACM.

[Chappell et al., 2015] Chappell, T., Geva, S., and Zuccon, G. (2015). Approximate nearest-neighbour search with inverted signature slice lists. In Hanbury, A., Kazai, G., Rauber, A., and Fuhr, N., editors, *Advances in Information Retrieval*, volume 9022 of *Lecture Notes in Computer Science*, pages 147–158. Springer International Publishing.

[Chathurani et al., 2015a] Chathurani, N., Geva, S., and V.Chandran (2015a). Conversion of an image to a document using grid-based decomposition for efficient content-based image retrieval. *International Journal of Information Science and Intelligent System*, 4.

[Chathurani et al., 2016a] Chathurani, N. W. U. D., Chappell, T., Geva, S., and Chandran, V. (2016a). Improving retrieval quality using pseudo relevance feedback in content-based image retrieval. In *39th Annual ACM Special Interest Group on Information Retrieval*.

[Chathurani et al., 2014] Chathurani, N. W. U. D., Geva, S., Chandran, V., and Chappell, T. (2014). Content based image retrieval using signature representation. In *12th Australasian Data Mining Conference (AusDM14)*.

[Chathurani et al., 2015b] Chathurani, N. W. U. D., Geva, S., Chandran, V., and Cynthujah, V. (2015b). An effective content based image retrieval system based on global representation and multi-level searching. In *2015 IEEE 10th International Conference on Industrial and Information Systems (ICIIS)*, pages 158–163.

[Chathurani et al., 2016b] Chathurani, N. W. U. D., Geva, S., Chandran, V., and Rajapaksha, P. (2016b). Image retrieval based on multi-feature fusion for heterogeneous image databases. In *18th International Conference on Image Analysis and Processing*.

[Chatzichristofis and Arampatzis, 2010] Chatzichristofis, S. A. and Arampatzis, A. (2010). Late fusion of compact composite descriptors for retrieval from heterogeneous image databases. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 825–826, New York, NY, USA. ACM.

[Chaudhary and Upadhyay, 2014] Chaudhary, M. D. and Upadhyay, A. B. (2014). Integrating shape and edge histogram descriptor with stationary wavelet transform for effective content based image retrieval. In *Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference on*, pages 1522–1527.

[Chen and Chandran, 2010] Chen, B. and Chandran, V. (2010). Robust image hashing using higher order spectral features. In *Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on*, pages 100–104.

[Chen and Wang, 2002] Chen, Y. and Wang, J. (2002). A region-based fuzzy feature matching approach to content-based image retrieval. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(9):1252–1267.

[Chen et al., 2003] Chen, Y., Wang, J., and Krovetz, R. (2003). An unsupervised learning approach to content-based image retrieval. In *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*, volume 1, pages 197–200 vol.1.

[Chen et al., 2005] Chen, Y., Wang, J., and Krovetz, R. (2005). Clue: cluster-based retrieval of images by unsupervised learning. *Image Processing, IEEE Transactions on*, 14(8):1187–1201.

[Chowdhury et al., 2012] Chowdhury, M., Das, S., and Kundu, M. (2012). Interactive content based image retrieval using ripplet transform and fuzzy relevance feedback. In Kundu, M., Mitra, S., Mazumdar, D., and Pal, S., editors, *Perception and Machine Intelligence*, volume 7143 of *Lecture Notes in Computer Science*, pages 243–251. Springer Berlin Heidelberg.

[Cox et al., 2000] Cox, I. J., Miller, M., Minka, T., Papathomas, T., and Yianilos, P. (2000). The bayesian image retrieval system, pichunter: theory, implementation, and psychophysical experiments. *Image Processing, IEEE Transactions on*, 9(1):20–37.

[Csurka et al., 2004] Csurka, G., Dance, C., Willamowski, J., Fan, L., and Bray, C. (2004). Visual categoriztion with bags of keypoints. In *ECCV International Workshop on Statistical Learning in Computer Vision*.

[Cui and Zhang, 2007] Cui, J. and Zhang, C. (2007). Combining stroke-based and selection-based relevance feedback for content-based image retrieval. In *Proceedings of the 15th International Conference on Multimedia*, MULTIMEDIA '07, pages 329–332, New York, NY, USA. ACM.

[Das, 2001] Das, S. (2001). Filters, wrappers and a boosting-based hybrid for feature selection. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 74–81, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[de Brito Ferreira, 2010] de Brito Ferreira, P. M. (2010). Content Űbased image classification: A non-parametric classification. Master's thesis, Universidade Tecnicia de Lisboa.

[De Vries and Geva, 2009a] De Vries, C. and Geva, S. (2009a). Document clustering with k-tree. In Geva, S., Kamps, J., and Trotman, A., editors, *Advances in Focused Retrieval*, volume 5631 of *Lecture Notes in Computer Science*, pages 420–431. Springer Berlin Heidelberg.

[De Vries and Geva, 2009b] De Vries, C. M. and Geva, S. (2009b). K-tree: Large scale document clustering. In *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '09, pages 718–719, New York, NY, USA. ACM.

[Dharani and Aroquiaraj, 2013] Dharani, T. and Aroquiaraj, I. L. (2013). A survey on content based image retrieval. In *Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on*, pages 485–490.

[Douik et al., 2016] Douik, A., Abdellaoui, M., and Kabbai, L. (2016). Content based image retrieval using local and global features descriptor. In *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pages 151–154.

[Egas et al., 1999] Egas, R., Huijsmans, N., Lew, M., and Sebe, N. (1999). *Adapting k-d Trees to Visual Retrieval*, pages 533–541. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Elharar et al., 2007] Elharar, E., Stern, A., Hadar, O., and Javidi, B. (2007). A hybrid compression method for integral images using discrete wavelet transform and discrete cosine transform. *Display Technology, Journal of*, 3(3):321–325.

[Estevez et al., 2009] Estevez, P. A., Tesmer, M., Perez, C. A., and Zurada, J. M. (2009). Normalized mutual information feature selection. *IEEE Transactions on Neural Networks*, 20(2):189–201.

[Estrada et al., 2004] Estrada, F., Jepson, A., and Fleet, D. (2004). Local features tutorial.

[Faloutsos and Christodoulakis, 1984] Faloutsos, C. and Christodoulakis, S. (1984). Signature files: An access method for documents and its analytical performance evaluation. *ACM Transactions on Information Systems (TOIS)*, 2(4):267–288.

[Fanhui, 2011] Fanhui, K. (2011). Image retrieval based on multi-features. In *Network Computing and Information Security (NCIS), 2011 International Conference on*, volume 1, pages 398–401.

[Farahat et al., 2013] Farahat, A. K., Ghodsi, A., and Kamel, M. S. (2013). Efficient greedy feature selection for unsupervised learning. *Knowledge and Information Systems*, 35(2):285–310.

[Fay and Proschan, 2010] Fay, M. P. and Proschan, M. A. (2010). Wilcoxon-mann-whitney or t-test? on assumptions for hypothesis tests and multiple interpretations of decision rules. *Statist. Surv.*, 4:1–39.

[Fei-Fei et al., 2004] Fei-Fei, L., Fergus, R., and Perona, P. (2004). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on*, pages 178–178.

[Feng et al., 2016] Feng, Y., Fan, L., and Wu, Y. (2016). Fast localization in large-scale environments using supervised indexing of binary features. *IEEE Transactions on Image Processing*, 25(1):343–358.

[Ferhatosmanoglu et al., 2001] Ferhatosmanoglu, H., Tuncel, E., Agrawal, D., and Abbadi, A. E. (2001). Approximate nearest neighbor searching in multimedia databases. In *Data Engineering, 2001. Proceedings. 17th International Conference on*, pages 503–511.

[Flickner et al., 1995] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., and Yanker, P. (1995). Query by image and video content: the qbic system. *Computer*, 28(9):23–32.

[Gallas et al., 2015] Gallas, A., Barhoumi, W., Kacem, N., and Zagrouba, E. (2015). Locality-sensitive hashing for region-based large-scale image indexing. *IET Image Processing*, 9(9):804–810.

[Gebara and Alhajj, 2007] Gebara, D. and Alhajj, R. (2007). Waveq: Combining wavelet analysis and clustering for effective image retrieval. In *Advanced Information Networking and Applications Workshops, 2007, AINAW '07. 21st International Conference on*, volume 1, pages 289–294.

[Geva, 2000] Geva, S. (2000). K-tree: a height balanced tree structured vector quantizer. In *Neural Networks for Signal Processing X, 2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop*, volume 1, pages 271–280 vol.1.

[Geva and De Vries, 2011] Geva, S. and De Vries, C. M. (2011). Topsig: Topology preserving document signatures. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, pages 333–338, New York, NY, USA. ACM.

[Gharsalli et al., 2015] Gharsalli, S., Emile, B., Laurent, H., Desquesnes, X., and Vivet, D. (2015). Random forest-based feature selection for emotion recognition. In *Image Processing Theory, Tools and Applications (IPTA), 2015 International Conference on*, pages 268–272.

[Gokalp and Aksoy, 2007] Gokalp, D. and Aksoy, S. (2007). Scene classification using bag-of-regions representations. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8.

[Goodrum, 2000] Goodrum, A. (2000). Image information retrieval: An overview of current research. *Informing Science*, 3:2000.

[Gordo et al., 2016] Gordo, A., Almazán, J., Revaud, J., and Larlus, D. (2016). *Deep Image Retrieval: Learning Global Representations for Image Search*, pages 241–257. Springer International Publishing, Cham.

[Gorman and Curran, 2006] Gorman, J. and Curran, J. R. (2006). Scaling distributional similarity to large corpora. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, ACL-44, pages 361–368, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Griffin et al., 2007] Griffin, G., Holub, A., and Perona, P. (2007). The caltech 256. Technical report, altech Technical Report.

[Gurrin et al., 2014] Gurrin, C., Smeaton, A. F., and Doherty, A. R. (2014). Lifelogging: Personal big data. *Foundations and Trends in Information Retrieval*, 8(1):1–125.

[Guttman, 1984] Guttman, A. (1984). R-trees: A dynamic index structure for spatial searching. In *Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data*, SIGMOD '84, pages 47–57, New York, NY, USA. ACM.

[Guyon and Elisseeff, 2003] Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research.*, 3:1157 – 1182.

[Harman, 1992] Harman, D. (1992). Relevance feedback revisited. In *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '92, pages 1–10, New York, NY, USA. ACM.

[Hays and Efros, 2008] Hays, J. and Efros, A. A. (2008). Scene completion using millions of photographs. *Commun. ACM*, 51(10):87–94.

[He et al., 2009] He, R., Zhu, Y., and Zhan, W. (2009). Using local latent semantic indexing with pseudo relevance feedback in web image retrieval. In *INC, IMS and IDC, 2009. NCM '09. Fifth International Joint Conference on*, pages 1354–1357.

[Heller and Ghahramani, 2006] Heller, K. and Ghahramani, Z. (2006). A simple bayesian framework for content-based image retrieval. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2110–2117.

[Hiremath and Pujari, 2007a] Hiremath, P. and Pujari, J. (2007a). Content based image retrieval using color, texture and shape features. In *Advanced Computing and Communications, 2007. ADCOM 2007. International Conference on*, pages 780–784.

[Hiremath and Pujari, 2007b] Hiremath, P. S. and Pujari, J. (2007b). Content based image retrieval based on color, texture and shape features using image and its complement. *International Journal of Computer Science and Security (IJCSS)*, 1:25–35.

[Hiremath and Pujari, 2008] Hiremath, P. S. and Pujari, J. (2008). Content based image retrieval using colour boosted salient points and shape features of an image. *Proc. International Journal of Image Processing*, 2:10–17.

[Hiwale et al., 2015] Hiwale, S. S., Dhotre, D., and Bamnote, G. R. (2015). Quick interactive image search in huge databases using content-based image retrieval. In *Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015 International Conference on*, pages 1–5.

[Hoi et al., 2008] Hoi, S. C. H., Liu, W., and Chang, S.-F. (2008). Semi-supervised distance metric learning for collaborative image retrieval. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–7.

[Howarth and Ruger, 2004] Howarth, P. and Ruger, S. (2004). Evaluation of texture features for content-based image retrieval. In *International Conference on Content-based Image and Video Retrieval (CIVR)*.

[Huang et al., 2006] Huang, X., Huang, Y. R., Wen, M., An, A., Liu, Y., and Poon, J. (2006). Applying data mining to pseudo-relevance feedback for high performance text retrieval. In *Sixth International Conference on Data Mining (ICDM'06)*, pages 295–306.

[Huang et al., 2016] Huang, Y., Zhu, F., Shao, L., and Frangi, A. F. (2016). Color object recognition via cross-domain learning on rgb-d images. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1672–1677.

[Huiskes and Lew, 2008] Huiskes, M. J. and Lew, M. S. (2008). The mir flickr retrieval evaluation. In *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval*, MIR '08, pages 39–43, New York, NY, USA. ACM.

[Hwang and Oh, 2015] Hwang, I. and Oh, S. (2015). A non-visual sensor triggered life logging system using canonical correlation analysis. In *Cyber-Physical Systems, Networks, and Applications (CPSNA), 2015 IEEE 3rd International Conference on*, pages 31–36.

[Iqbal and Aggarwal, 2000] Iqbal, Q. and Aggarwal, J. (2000). Lower-level and higher-level approaches to content-based image retrieval. In *Image Analysis and Interpretation, 2000. Proceedings. 4th IEEE Southwest Symposium*, pages 197–201.

[Iqbal and Aggarwal, 2003] Iqbal, Q. and Aggarwal, J. (2003). Feature integration, multi-image queries and relevance feedback in image retrieval. In *6th International Conference on Visual Information Systems (VISUAL)*, pages 467–474.

[Ishikawa et al., 1998] Ishikawa, Y., Subramanya, R., and Faloutsos, C. (1998). Mindreader: Querying databases through multiple examples. *Computer Science Department*, page 551.

[Jain, 2010] Jain, M. (2010). Towards efficient and scalable visual processing in images and videos. master thesis. Master's thesis, International Institute of Information Technology.

[Javidi et al., 2008] Javidi, M., Aski, B., Homaei, H., and Pourreza, H. (2008). A new approach for interactive image retrieval based on fuzzy feedback and support vector machine. In *Computational Intelligence for Modelling Control Automation, 2008 International Conference on*, pages 1205–1210.

[Jiang et al., 2006] Jiang, W., Er, G., Dai, Q., and Gu, J. (2006). Similarity-based online feature selection in content-based image retrieval. *Image Processing, IEEE Transactions on*, 15(3):702–712.

[Jiang et al., 2016] Jiang, Y. G., Wang, J., Wang, Q., Liu, W., and Ngo, C. W. (2016). Hierarchical visualization of video search results for topic-based browsing. *IEEE Transactions on Multimedia*, PP(99):1–1.

[Jing et al., 2002] Jing, F., Li, M., Zhang, H.-J., and Zhang, B. (2002). An effective region-based image retrieval framework. In *Proceedings of the Tenth ACM International Conference on Multimedia*, MULTIMEDIA '02, pages 456–465, New York, NY, USA. ACM.

[Jovic et al., 2006] Jovic, M., Hatakeyama, Y., Dong, F., and Hirota, K. (2006). Image retrieval based on similarity score fusion from feature similarity ranking lists. In Wang, L., Jiao, L., Shi, G., Li, X., and Liu, J., editors, *Fuzzy Systems and Knowledge Discovery*, volume 4223 of *Lecture Notes in Computer Science*, pages 461–470. Springer Berlin Heidelberg.

[Karpathy and Fei-Fei, 2015] Karpathy, A. and Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3128–3137.

[Kato, 1992] Kato, T. (1992). Database architecture for content-based image retrieval. In *Proc. SPIE 1662, Image Storage and Retrieval Systems*.

[Khavare and Manjrekar, 2015] Khavare, S. A. and Manjrekar, A. A. (2015). Robust image hashing algorithm for detecting and localizing image tamper in small region. In *2015 International Conference on Information Processing (ICIP)*, pages 289–294.

[Khelifi and Jiang, 2010] Khelifi, F. and Jiang, J. (2010). Perceptual image hashing based on virtual watermark detection. *IEEE Transactions on Image Processing*, 19(4):981–994.

[Kherfi and Ziou, 2006] Kherfi, M. and Ziou, D. (2006). Relevance feedback for cbir: a new approach based on probabilistic feature weighting with positive and negative examples. *Image Processing, IEEE Transactions on*, 15(4):1017–1030.

[Kidiyo and Joseph, 2008] Kidiyo, Y. M. K. and Joseph, R. (2008). A survey of shape feature extraction techniques.

[Ko and Byun, 2002] Ko, B. and Byun, H. (2002). Probabilistic neural networks supporting multi-class relevance feedback in region-based image retrieval. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 4, pages 138–141 vol.4.

[Kogler and Lux, 2010] Kogler, M. and Lux, M. (2010). Bag of visual words revisited: An exploratory study on robust image retrieval exploiting fuzzy codebooks. In *Proceedings of the Tenth International Workshop on Multimedia Data Mining*, MDMKDD '10, pages 3:1–3:6, New York, NY, USA. ACM.

[Kohavi and John, 1997] Kohavi, R. and John, G. H. (1997). Wrappers for feature subset selection. *Artif. Intell.*, 97(1-2):273–324.

[Kokare et al., 2005] Kokare, M., Biswas, P., and Chatterji, B. (2005). Texture image retrieval using new rotated complex wavelet filters. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 35(6):1168–1178.

[Laaksonen et al., 2002] Laaksonen, J., Koskela, M., and Oja, E. (2002). Picsom-self-organizing image retrieval with mpeg-7 content descriptors. *Neural Networks, IEEE Transactions on*, 13(4):841–853.

[Lazebnik et al., 2006] Lazebnik, S., Schmid, C., and Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2169–2178.

[Lazebnik et al., 2009] Lazebnik, S., Schmid, C., and Ponce, J. (2009). Spatial pyramid matching. In *Object Categorization: Computer and Human Vision Perspectives.*

[Lew et al., 2006] Lew, M. S., Sebe, N., Djeraba, C., and Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput. Commun. Appl.*, 2(1):1–19.

[Li et al., 2006] Li, J., Allinson, N., Tao, D., and Li, X. (2006). Multitraining support vector machine for image retrieval. *Image Processing, IEEE Transactions on*, 15(11):3597–3601.

[Li and Allinson, 2013] Li, J. and Allinson, N. M. (2013). Relevance feedback in content-based image retrieval: a survey. In *Handbook on neural information processing*, pages 433–469. Springer International Publishing.

[Li and Wang, 2003] Li, J. and Wang, J. (2003). Automatic linguistic indexing of pictures by a statistical modeling approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1075–1088.

[Li et al., 2000] Li, J., Wang, J. Z., and Wiederhold, G. (2000). Irm: Integrated region matching for image retrieval. In *Proceedings of the Eighth ACM International Conference on Multimedia*, MULTIMEDIA '00, pages 147–156, New York, NY, USA. ACM.

[Li et al., 2008a] Li, M., Ng, M., ming Cheung, Y., and Huang, J. (2008a). Agglomerative fuzzy k-means clustering algorithm with selection of number of clusters. *Knowledge and Data Engineering, IEEE Transactions on*, 20(11):1519–1534.

[Li et al., 2016] Li, S., Lum, A., Brahm, G., Nachum, I. B., Sharma, M., Shmuilovich, O., and Warrington, J. (2016). A bag-of-shapes descriptor for medical imaging. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2316–2320.

[Li et al., 2008b] Li, X., Wu, C., Zach, C., Lazebnik, S., and Frahm, J.-M. (2008b). Modeling and recognition of landmark image collections using iconic scene graphs. In *Proceedings of the 10th European Conference on Computer Vision: Part I*, ECCV '08, pages 427–440, Berlin, Heidelberg. Springer-Verlag.

[Li and Wang, 2016] Li, Y. and Wang, P. (2016). Robust image hashing based on low-rank and sparse decomposition. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2154–2158.

[Lim et al., 2016] Lim, C. D., Wang, C. M., Cheng, C. Y., Chao, Y., Tseng, S. H., and Fu, L. C. (2016). Sensory cues guided rehabilitation robotic walker realized by depth image-based gait analysis. *IEEE Transactions on Automation Science and Engineering*, 13(1):171–180.

[Lin et al., 2009] Lin, C.-H., Chen, R.-T., and Chan, Y.-K. (2009). A smart content-based image retrieval system based on color and texture feature. *Image and Vision Computing*, 27(6):658 – 665.

[Lin et al., 2003] Lin, W.-H., Jin, R., and Hauptmann, A. (2003). Web image retrieval re-ranking with relevance model. In *Proceedings of the 2003 IEEE/WIC International Conference on Web Intelligence*, WI '03, pages 242–, Washington, DC, USA. IEEE Computer Society.

[Liu and Yu, 2005] Liu, H. and Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17(4):491–502.

[Liu et al., 2013] Liu, J., Danait, N., Hu, S., and Sengupta, S. (2013). A leave-one-feature-out wrapper method for feature selection in data classification. In *2013 6th International Conference on Biomedical Engineering and Informatics*, pages 656–660.

[Liu et al., 2008] Liu, Y., Zhang, D., and Lu, G. (2008). Region-based image retrieval with high-level semantics using decision tree learning. *Pattern Recognition*, 41(8):2554 – 2570.

[Liu et al., 2007] Liu, Y., Zhang, D., Lu, G., and Ma, W.-Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1):262 – 282.

[Liu et al., 2014] Liu, Z., Li, H., Zhou, W., Zhao, R., and Tian, Q. (2014). Contextual hashing for large-scale image search. *IEEE Transactions on Image Processing*, 23(4):1606–1614.

[Liua et al., 2007] Liua, Y., Zhanga, D., Lua, G., and Mab, W. (2007). A survey of content-based image retrieval with high-level semantics. *Journal of Pattern Recognition*, 40:262–282.

[Long et al., 2003] Long, F., Zhang, H., and Feng, D. D. (2003). *Fundamentals of Content-Based Image Retrieval*, pages 1–26. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Loorak et al., 2016] Loorak, M. H., Perin, C., Kamal, N., Hill, M., and Carpendale, S. (2016). Timespan: Using visualization to explore temporal multi-dimensional data of stroke patients. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):409–418.

[Lu et al., 2000] Lu, Y., Hu, C., Zhu, X., Zhang, H., and Yang, Q. (2000). A unified framework for semantics and feature based relevance feedback in image retrieval systems. In *Proceedings of the eighth ACM international conference on Multimedia*, pages 31–37. ACM.

[Lv and Wang, 2013] Lv, X. and Wang, Z. J. (2013). Compressed binary image hashes based on semisupervised spectral embedding. *IEEE Transactions on Information Forensics and Security*, 8(11):1838–1849.

[Ma and Manjunath, 1997] Ma, W. and Manjunath, B. (1997). Netra: a toolbox for navigating large image databases. In *Image Processing, 1997. Proceedings., International Conference on*, volume 1, pages 568–571 vol.1.

[MacQueen, 1967] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, Berkeley, Calif. University of California Press.

[Maldonado and Weber, 2009] Maldonado, S. and Weber, R. (2009). A wrapper method for feature selection using support vector machines. *Information Sciences*, 179(13):2208 – 2217. Special Section on High Order Fuzzy Sets.

[Manjunath et al., 2001] Manjunath, B. S., Ohm, J. R., Vasudevan, V. V., and Yamada, A. (2001). Color and texture descriptors. *IEEE Trans. Cir. and Sys. for Video Technol.*, 11(6):703–715.

[Manning et al., 2008] Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.

[Mansoori et al., 2013] Mansoori, N., Nejati, M., Razzaghi, P., and Samavi, S. (2013). Bag of visual words approach for image retrieval using color information. In *Electrical Engineering (ICEE), 2013 21st Iranian Conference on*, pages 1–6.

[Manthalkar et al., 2003] Manthalkar, R., Biswas, P., and Chatterji, B. (2003). Rotation and scale invariant texture features using discrete wavelet packet transform. *Pattern Recognition Letters*, 24(14):2455 – 2462.

[Mariam and R, 2015] Mariam, N. and R, R. (2015). A modified approach in cbir based on combined edge detection, color and discrete wavelet transform. In *Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on*, pages 2201–2205.

[Mehrotra et al., 1997] Mehrotra, S., Rui, Y., Chakrabarti, K., Ortega, M., and Huang, T. S. (1997). Multimedia analysis and retrieval system. In *Proceeding of the 3rd international workshop on Information Retrieval Systems*.

[Moffat and Zobel, 1996] Moffat, A. and Zobel, J. (1996). Self-indexing inverted files for fast text retrieval. *ACM Trans. Inf. Syst.*, 14(4):349–379.

[Murata et al., 2015] Murata, R., Mishina, Y., Yamauchi, Y., Yamashita, T., and Fujiyoshi, H. (2015). Efficient feature selection method using contribution ratio by random forest. In *Frontiers of Computer Vision (FCV), 2015 21st Korea-Japan Joint Workshop on*, pages 1–6.

[Murthy et al., 2010] Murthy, V. S. V. S., Vamsidhar, E., Swarup Kumar, J. N. V. R., and Sankara Rao, P. (2010). Content based image retrieval using hierarchical and k-means clustering techniques. *International Journal of Engineering Science and Technology*, 2:209–212.

[Nguyen et al., 2016] Nguyen, H. T., Jung, S. W., and Won, C. S. (2016). Order-preserving condensation of moving objects in surveillance videos. *IEEE Transactions on Intelligent Transportation Systems*, 17(9):2408–2418.

[Niblack et al., 1993] Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, E., Petkovic, D., and Yanker, P. (1993). The qbic project: Querying image by content using colour,texture,shape. In *In Proceeding SPIE Storage and Retrieval for Image and Video Databases*, volume 1908, pages 173–187.

[Nister and Stewenius, 2006] Nister, D. and Stewenius, H. (2006). Scalable recognition with a vocabulary tree. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, CVPR '06, pages 2161–2168, Washington, DC, USA. IEEE Computer Society.

[Noteboom et al., 2014] Noteboom, C. B., Motorny, S. P., Qureshi, S., and Sarnikar, S. (2014). Meaningful use of electronic health records for physician collaboration: A patient centered health care perspective. In *2014 47th Hawaii International Conference on System Sciences*, pages 656–666.

[O'hara and Draper, 2010] O'hara, S. and Draper, B. A. (2010). Introduction to the bag of features paradigm for image classification and retrieval.

[Oliva and Torralba, 2001] Oliva, A. and Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vision*, 42(3):145–175.

[Oussalah, 2008] Oussalah, M. (2008). Content based image retrieval: Review of state of art and future directions. In *2008 First Workshops on Image Processing Theory, Tools and Applications*, pages 1–10.

[Palermo and Weller, 1980] Palermo, F. P. and Weller, D. (1980). Some database requirements for pictorial applications. In *Proceedings on Data Base Techniques for Pictorial Applications*, pages 555–567, London, UK, UK. Springer-Verlag.

[Pass et al., 1996] Pass, G., Zabih, R., and Miller, J. (1996). Comparing images using color coherence vectors. In *Proceedings of the Fourth ACM International Conference on Multimedia*, MULTIMEDIA '96, pages 65–73, New York, NY, USA. ACM.

[Patvardhan et al., 2013] Patvardhan, C., Verma, A. K., and Lakshmi, C. V. (2013). Robust content based image retrieval based on multi-resolution wavelet features and edge histogram. In *Image Information Processing (ICIIP), 2013 IEEE Second International Conference on*, pages 447–452.

[Peker, 2011] Peker, K. (2011). Binary sift: Fast image retrieval using binary quantized sift features. In *Content-Based Multimedia Indexing (CBMI), 2011 9th International Workshop on*, pages 217–222.

[Peng et al., 2005] Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238.

[Pentland et al., 1996] Pentland, A., Picard, R. W., and Sclaroff, S. (1996). Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3):233–254.

[Philbin et al., 2008] Philbin, J., Chum, O., Isard, M., Sivic, J., and Zisserman, A. (2008). Lost in quantization: Improving particular object retrieval in large scale image databases. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8.

[Pichler et al., 1996] Pichler, O., Teuner, A., and Hosticka, B. J. (1996). A comparison of texture feature extraction using adaptive gabor filtering, pyramidal and tree structured wavelet transforms. *Pattern Recogn.*, 29(5):733–742.

[Prasad and Leung, 2010] Prasad, D. K. and Leung, M. K. H. (2010). An ellipse detection method for real images. In *Image and Vision Computing New Zealand (IVCNZ), 2010 25th International Conference of*, pages 1–8.

[Qian et al., 2016] Qian, X., Tan, X., Zhang, Y., Hong, R., and Wang, M. (2016). Enhancing sketch-based image retrieval by re-ranking and relevance feedback. *IEEE Transactions on Image Processing*, 25(1):195–208.

[Qiu, 2002] Qiu, G. (2002). Indexing chromatic and achromatic patterns for content-based colour image retrieval. *Pattern Recognition*, 35(8):1675 – 1686. Colour Imaging.

[Rahat et al., 2012] Rahat, K., Cecile, B., Damien, M., and Christophe, D. (2012). Spatial orientations of visual word pairs to improve bag-of-visual-words model. In *British Machine Vision Conference*, pages 1–11.

[Rahman et al., 2007] Rahman, M., Bhattacharya, P., and Desai, B. (2007). A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. *Information Technology in Biomedicine, IEEE Transactions on*, 11(1):58–69.

[Rahman et al., 2006] Rahman, M., Desai, B., and Bhattacharya, P. (2006). A feature level fusion in similarity matching to content-based image retrieval. In *Information Fusion, 2006 9th International Conference on*, pages 1–6.

[Rahmana et al., 2011] Rahmana, M., Pickering, M., and Frater, M. (2011). Scale and rotation invariant gabor features for texture retrieval. In *Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on*, pages 602–607.

[Rai et al., 2011] Rai, H., Shen, X., Deepak, K., and Krishna, P. (2011). Hybrid feature to encode shape and texture for content based image retrieval. In *Image Information Processing (ICIIP), 2011 International Conference on*, pages 1–6.

[Rocchio, 1971] Rocchio, J. (1971). Relevance feedback in information retrieval. *THE SMART RETRIEVAL SYSTEM: Experiments in Automatic Document Processing*, pages 313–323.

[Roth, 2004] Roth, V. (2004). The generalized lasso. *IEEE Transactions on Neural Networks*, 15(1):16–28.

[Rudinac et al., 2009] Rudinac, S., Larson, M., and Hanjalic, A. (2009). Exploiting visual reranking to improve pseudo-relevance feedback for spoken-content-based video retrieval. In *2009 10th Workshop on Image Analysis for Multimedia Interactive Services*, pages 17–20.

[Rui et al., 1997] Rui, Y., Huang, T., and Mehrotra, S. (1997). Content-based image retrieval with relevance feedback in mars. In *Image Processing, 1997. Proceedings., International Conference on*, volume 2, pages 815–818 vol.2.

[Rui et al., 1999] Rui, Y., Huang, T. S., and Chang, S.-F. (1999). Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, 10(1):39 – 62.

[Ruikar and Kabade, 2016] Ruikar, S. D. and Kabade, R. S. (2016). Content based image retrieval by combining feature vector. In *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pages 1517–1523.

[Saad et al., 2011] Saad, M., Saleh, H., Konbor, H., and Ashour, M. (2011). Image retrieval based on integration between ycbcr color histogram and shape feature. In *Computer Engineering Conference (ICENCO), 2011 Seventh International*, pages 97–102.

[Saavedra, 2014] Saavedra, J. M. (2014). Sketch based image retrieval using a soft computation of the histogram of edge local orientations (s-helo). In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 2998–3002.

[Saha et al., 2007] Saha, S. K., Das, A. K., and Chanda, B. (2007). Image retrieval based on indexing and relevance feedback. *Pattern Recognition Letters*, 28(3):357 – 366. Advances in Visual information ProcessingSpecial Issue of Pattern Recognition Letters on Advances in Visual Information Processing. (ICVGIP 2004).

[Sahlgren, 2005] Sahlgren, M. (2005). An introduction to random indexing. In *In Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering, TKE 2005*.

[Salton et al., 1975] Salton, G., Wong, A., and Yang, C. S. (1975). A vector space model for automatic indexing. *Commun. ACM*, 18(11):613–620.

[Sharma et al., 2011] Sharma, N., Rawat, P., and Singh, J. (2011). Efficient cbir using color histogram processing. signal & image processing : An international jouranal (sipij). *n International Jouranal (SIPIJ)*, 2:94–112.

[Shaw, 1995] Shaw, W. J. (1995). Term-relevance computations and perfect retrieval performance. *Information Processing & Management*, 31(4):491 – 498.

[Shrivastava et al., 2015] Shrivastava, S., Gupta, B., and Gupta, M. (2015). Optimization of image retrieval by using hsv color space, zernike moment dwt technique. In *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pages 1–5.

[Shu et al., 2015] Shu, J., Hu, A., Meng, F., and Meng, Y. (2015). Improved binary codes for efficient content-based image retrieval. In *2015 8th International Congress on Image and Signal Processing (CISP)*, pages 550–554.

[Silva et al., 2013] Silva, F. B., Goldenstein, S., Tabbone, S., and da S. Torres, R. (2013). Image classification based on bag of visual graphs. In *2013 IEEE International Conference on Image Processing*, pages 4312–4316.

[Sivic and Zisserman, 2003] Sivic, J. and Zisserman, A. (2003). Video google: a text retrieval approach to object matching in videos. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1470–1477 vol.2.

[Smeulders et al., 2000] Smeulders, A., Worring, M., Santini, S., Gupta, A., and Jain, R. (2000). Content-based image retrieval at the end of the early years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(12):1349–1380.

[Smith and Chang, 1996] Smith, J. R. and Chang, S.-F. (1996). Visualseek: A fully automated content-based image query system. In *Proceedings of the Fourth ACM International Conference on Multimedia*, MULTIMEDIA '96, pages 87–98, New York, NY, USA. ACM.

[Smith and fu Chang, 1996a] Smith, J. R. and fu Chang, S. (1996a). Querying by color regions using the visualseek content-based visual query system. In *in Intelligent Multimedia Information Retrieval*, pages 23–41. AAAI Press.

[Smith and fu Chang, 1996b] Smith, J. R. and fu Chang, S. (1996b). Tools and techniques for color image retrieval. In *Storage & Retrieval for Image and Video Databases*, pages 426–437.

[Squire et al., 1999] Squire, D., Müller, W., and Müller, H. (1999). *Relevance Feedback and Term Weighting Schemes for Content-Based Image Retrieval*, pages 549–557. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Su et al., 2011a] Su, J. H., Huang, W. J., Yu, P. S., and Tseng, V. S. (2011a). Efficient relevance feedback for content-based image retrieval by mining user navigation patterns. *IEEE Transactions on Knowledge and Data Engineering*, 23(3):360–372.

[Su et al., 2011b] Su, J. H., Huang, W. J., Yu, P. S., and Tseng, V. S. (2011b). Efficient relevance feedback for content-based image retrieval by mining user navigation patterns. *IEEE Transactions on Knowledge and Data Engineering*, 23(3):360–372.

[Su et al., 2003] Su, Z., Zhang, H., Li, S., and Ma, S. (2003). Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning. *Image Processing, IEEE Transactions on*, 12(8):924–937.

[Tahoun et al., 2005] Tahoun, M., Nagaty, K., El-Arief, T., and A-Megeed, M. (2005). A robust content-based image retrieval system using multiple features representations. In *Networking, Sensing and Control, 2005. Proceedings. 2005 IEEE*, pages 116–122.

[Takahashi et al., 2000] Takahashi, N., Iwasaki, M., Kunieda, T., Wakita, Y., and Day, N. (2000). Image retrieval using spatial intensity features. *Signal Processing: Image Communication*, 16(1âĂŞ2):45 – 57.

[Takala et al., 2005] Takala, V., Ahonen, T., and PietikÃd'inen, M. (2005). Block-based methods for image retrieval using local binary patterns. In Kalviainen, H., Parkkinen, J., and Kaarna, A., editors, *Image Analysis*, volume 3540 of *Lecture Notes in Computer Science*, pages 882–891. Springer Berlin Heidelberg.

[Tang et al., 2016] Tang, Z., Zhang, X., Li, X., and Zhang, S. (2016). Robust image hashing with ring partition and invariant vector distance. *IEEE Transactions on Information Forensics and Security*, 11(1):200–214.

[Tao et al., 2007] Tao, D., Li, X., and Maybank, S. (2007). Negative samples analysis in relevance feedback. *Knowledge and Data Engineering, IEEE Transactions on*, 19(4):568–580.

[Tao et al., 2006] Tao, D., Tang, X., Li, X., and Rui, Y. (2006). Direct kernel biased discriminant analysis: a new content-based image retrieval relevance feedback algorithm. *Multimedia, IEEE Transactions on*, 8(4):716–727.

[Torralba et al., 2008a] Torralba, A., Fergus, R., and Freeman, W. (2008a). 80 million tiny images: A large data set for nonparametric object and scene recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(11):1958–1970.

[Torralba et al., 2008b] Torralba, A., Fergus, R., and Weiss, Y. (2008b). Small codes and large image databases for recognition. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8.

[Veltkamp, 2001] Veltkamp, R. C. (2001). Shape matching: similarity measures and algorithms. In *Shape Modeling and Applications, SMI 2001 International Conference on.*, pages 188–197.

[Vries et al., 2009] Vries, C. M. D., V., V. L. D., and G., S. (2009). Random indexing k-tree. In *Proc. ADCS of the 14th Australian Document Computing Symposium.*

[Wan et al., 2014] Wan, J., Wang, D., Hoi, S. C. H., Wu, P., Zhu, J., Zhang, Y., and Li, J. (2014). Deep learning for content-based image retrieval: A comprehensive study. In *Proceedings of the 22Nd ACM International Conference on Multimedia*, MM '14, pages 157–166, New York, NY, USA. ACM.

[Wang et al., 2014] Wang, A., An, N., Chen, G., Yang, J., Li, L., and Alterovitz, G. (2014). Incremental wrapper based gene selection with markov blanket. In *Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on*, pages 74–79.

[Wang et al., 2006] Wang, B., Zhang, X., and Li, N. (2006). Relevance feedback technique for content-based image retrieval using neural network learning. In *Machine Learning and Cybernetics, 2006 International Conference on*, pages 3692–3696.

[Wang et al., 2001] Wang, J., Li, J., and Wiederhold, G. (2001). Simplicity: semantics-sensitive integrated matching for picture libraries. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(9):947–963.

[Wang et al., 2008] Wang, Q., Ye, Y., and Huang, J. (2008). Fuzzy k-means with variable weighting in high dimensional data analysis. In *Web-Age Information Management, 2008. WAIM '08. The Ninth International Conference on*, pages 365–372.

[Wang et al., 2015] Wang, X., Pang, K., Zhou, X., Zhou, Y., Li, L., and Xue, J. (2015). A visual model-based perceptual image hash for content authentication. *IEEE Transactions on Information Forensics and Security*, 10(7):1336–1349.

[Wu and Yap, 2006] Wu, K. and Yap, K.-H. (2006). Fuzzy svm for content-based image retrieval: a pseudo-label support vector machine framework. *Computational Intelligence Magazine, IEEE*, 1(2):10–16.

[Xiao et al., 2010] Xiao, J., Hays, J., Ehinger, K., Oliva, A., and Torralba, A. (2010). Sun database: Large-scale scene recognition from abbey to zoo. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3485–3492.

[Xie et al., 2013] Xie, W., Lu, Z., Peng, Y., and Xiao, J. (2013). Multimodal semi-supervised image classification by combining tag refinement, graph-based learning and support vector regression. In *2013 IEEE International Conference on Image Processing*, pages 4307–4311.

[Xu et al., 2007] Xu, D., Yan, S., Tao, D., Lin, S., and Zhang, H.-J. (2007). Marginal fisher analysis and its variants for human gait recognition and content-based image retrieval. *Image Processing, IEEE Transactions on*, 16(11):2811–2821.

[Yan et al., 2003] Yan, R., Hauptmann, A. G., and Jin, R. (2003). Negative pseudo-relevance feedback in content-based video retrieval. In *Proceedings of the Eleventh ACM International Conference on Multimedia*, MULTIMEDIA '03, pages 343–346, New York, NY, USA. ACM.

[Yang et al., 2007] Yang, J., Jiang, Y.-G., Hauptmann, A. G., and Ngo, C.-W. (2007). Evaluating bag-of-visual-words representations in scene classification. In *Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval*, MIR '07, pages 197–206, New York, NY, USA. ACM.

[Yang et al., 2009] Yang, J., Yu, K., Gong, Y., and Huang, T. (2009). Linear spatial pyramid matching using sparse coding for image classification. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1794–1801.

[Yang et al., 2015a] Yang, X., Lv, F., Cai, L., and Li, D. (2015a). Adaptive learning region importance for region-based image retrieval. *IET Computer Vision*, 9(3):368–377.

[Yang et al., 2015b] Yang, Y., Shen, F., Shen, H. T., Li, H., and Li, X. (2015b). Robust discrete spectral hashing for large-scale image semantic indexing. *IEEE Transactions on Big Data*, 1(4):162–171.

[Yap and Wu, 2007] Yap, K.-H. and Wu, K. (2007). A pseudo-labeling framework for content-based image retrieval. In *Computational Intelligence in Image and Signal Processing, 2007. CIISP 2007. IEEE Symposium on*, pages 266–270.

[yi Lee and Lee, 2013] yi Lee, H. and Lee, L.-S. (2013). Enhanced spoken term detection using support vector machines and weighted pseudo examples. *Audio, Speech, and Language Processing, IEEE Transactions on*, 21(6):1272–1284.

[Yoon et al., 2014] Yoon, J., Choi, J., and Yoo, C. D. (2014). A hierarchical-structured dictionary learning for image classification. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 155–159.

[Yu and Yang, 2001] Yu, H. and Yang, J. (2001). A direct lda algorithm for high-dimensional data—with application to face recognition. *Pattern recognition*, 34(10):2067–2070.

[Yu and Liu, 2003] Yu, L. and Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the Twentieth International Conference on Machine Learning*, pages 856–863.

[Yuan et al., 2013a] Yuan, L., Liu, J., and Ye, J. (2013a). Efficient methods for overlapping group lasso. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9):2104–2116.

[Yuan et al., 2011a] Yuan, P. H., Yang, K. F., and Tsai, W. H. (2011a). Real-time security monitoring around a video surveillance vehicle with a pair of two-camera omni-imaging devices. *IEEE Transactions on Vehicular Technology*, 60(8):3603–3614.

[Yuan et al., 2011b] Yuan, X. Y., J. Qin, Z., and Wan, T. (2011b). A sift-lbp image retrieval model based on bag-of-features. In *Proc. IEEE ICIP 18th International Conference on Image Processing*, pages 1061–1064.

[Yuan et al., 2013b] Yuan, Z., Sang, J., and Xu, C. (2013b). Tag-aware image classification via nested deep belief nets. In *2013 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6.

[Zepeda et al., 2009] Zepeda, J., Kijak, E., and Guillemot, C. (2009). Sift-based local image description using sparse representations. In *Multimedia Signal Processing, 2009. MMSP '09. IEEE International Workshop on*, pages 1–6.

[Zhang et al., 2009] Zhang, D., Islam, M. M., Lu, G., and Hou, J. (2009). Semantic image retrieval using region based inverted file. In *Digital Image Computing: Techniques and Applications, 2009. DICTA '09.*, pages 242–249.

[Zhang and Lu, 2002] Zhang, D. and Lu, G. (2002). Shape-based image retrieval using generic fourier descriptor. *Signal Processing: Image Communication*, 17(10):825 – 848.

[Zhang and Ye, 2009] Zhang, J. and Ye, L. (2009). Watermarking protocol for protecting user's right in content based image retrieval. In *2009 IEEE International Conference on Multimedia and Expo*, pages 1082–1085.

[Zhang et al., 2012a] Zhang, L., Wang, L., and Lin, W. (2012a). Conjunctive patches subspace learning with side information for collaborative image retrieval. *IEEE Transactions on Image Processing*, 21(8):3707–3720.

[Zhang et al., 2012b] Zhang, L., Wang, L., and Lin, W. (2012b). Generalized biased discriminant analysis for content-based image retrieval. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(1):282–290.

[Zhang et al., 2012c] Zhang, L., Wang, L., and Lin, W. (2012c). Semisupervised biased maximum margin analysis for interactive image retrieval. *IEEE Transactions on Image Processing*, 21(4):2294–2308.

[Zhang et al., 2014] Zhang, L., Wang, L., Lin, W., and Yan, S. (2014). Geometric optimum experimental design for collaborative image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(2):346–359.

[Zhao et al., 2013a] Zhao, Y., Wang, S., Zhang, X., and Yao, H. (2013a). Robust hashing for image authentication using zernike moments and local features. *IEEE Transactions on Information Forensics and Security*, 8(1):55–63.

[Zhao et al., 2013b] Zhao, Z., Wang, L., Liu, H., and Ye, J. (2013b). On similarity preserving feature selection. *IEEE Trans. on Knowl. and Data Eng.*, 25(3):619–632.

[Zhen-Sheng, 2012] Zhen-Sheng, N. (2012). B-sift: A binary sift based local image feature descriptor. In *Digital Home (ICDH), 2012 Fourth International Conference on*, pages 117–121.

[Zheng et al., 2006] Zheng, Q.-F., Wang, W.-Q., and Gao, W. (2006). Effective and efficient object-based image retrieval using visual phrases. In *Proceedings of the 14th Annual ACM International Conference on Multimedia*, MULTIMEDIA '06, pages 77–80, New York, NY, USA. ACM.

[Zhou and Huang, 2001a] Zhou, X. S. and Huang, T. S. (2001a). Comparing discriminating transformations and svm for learning during multimedia retrieval. In *Proceedings of the ninth ACM international conference on Multimedia*, pages 137–146. ACM.

[Zhou and Huang, 2001b] Zhou, X. S. and Huang, T. S. (2001b). Small sample learning during multimedia retrieval using biasmap. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–11. IEEE.

[Zou and Hastie, 2005] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Statist. Soc. B*, 67:301Ű320.